Master Thesis

# ENRICHING A QUESTION-ANSWERING SYSTEM WITH USER EXPERIENCE CONCEPTS

May 2019

Utrecht University

Department of Information and Computing Sciences
Master Business Informatics

| | | | |
|---|---|---|---|
| Author: | L.V. (Roy) Simons, BSc | 1st Supervisor: | dr. M.R. (Marco) Spruit |
| Student Number: | 6028527 | 2nd Supervisor: | dr. L.M. (Marcela) Ruiz Carmona |
| | | Daily Supervisor: | ir. J.R. (Rob) Douwes |

Version: 1.1

# ACKNOWLEDGEMENTS

First, I would like to thank Marco Spruit for his support during the thesis process as my first supervisor, and I would like to thank Marcela Ruiz for the additional support in her role as my second supervisor. I would like to thank Rob Douwes for his extraordinary support and enthusiasm during the internship as my daily supervisor. His feedback and insights motivated me to work harder and in a more creative manner. He continuously challenged me to broaden my horizons. I would also like to thank him and his team for making this thesis process a fun experience as well besides all seriousness. The same goes for the fellow interns with who I bonded more than I could have imagined. A special thanks goes out to my girlfriend, Marceline, who was always there for me when I needed it the most. Lastly, I would like to thank my family and friends for their endless support.

# ABSTRACT

While question-answering systems are extensively researched, its user experience has not been investigated yet. Therefore, it remains unclear how a question-answering system can be enriched with user experience concepts. ExperienceBank wants to automate question-answering, and while a question-answering system could be utilized to answer frequently asked customer questions automatically, it is unclear how their desired '9+ customer experience' can then still be achieved. Therefore, the present research is conducted at ExperienceBank to clarify how a question-answering system can be enriched with user experience concepts to better handle frequently asked customer questions. A literature review and an empirical investigation have been conducted to explore the body of knowledge on question-answering systems and user experience. This led to theoretical and empirical knowledge that clarified the need for and constituted the foundation for the *user experience enrichment guidelines for question-answering systems*. These guidelines have been designed by prescribing how to (1) manage user anticipations, (2) address human needs, and (3) measure user opinions, as these areas have been found to be most important in enabling positive user experience. A validation by expert opinions has been conducted to gather qualitative data on the adequacy and shortcomings of the guidelines. This showed that although the guidelines enable enrichment of a question-answering system with useful and effective user experience concepts, they need to be optimized. Several improvements were suggested but the main issue is the lack of instruction on balancing the enrichments, causing a too excessively enriched question-answering system. This is predicted to have a counterproductive effect on its usage and experience. More research on the effectiveness of particular (combinations of) enrichments must be conducted to determine a prioritization of enrichments and eventually an enrichment optimum.

# TABLE OF CONTENTS

# LIST OF FIGURES AND TABLES

# ABBREVIATIONS

| Abbreviation | Explanation |
|---|---|
| CSR | Customer service representative |
| HCI | Human-computer interaction |
| IE | Information extraction |
| IR | Information retrieval |
| NLP | Natural language processing |
| NLU | Natural language understanding |
| PoC | Proof of concept |
| QAS | Question-answering system |
| SST | Self-service technology |
| UX | User experience |
| XBank | ExperienceBank |

# 1    INTRODUCTION

In this chapter, the background of this research that is conducted for the master thesis is described by introducing the central subjects and the research gap for which this research should produce scientific value. This research is purposed to define how a question-answering system (QAS) can be enriched with user experience (UX) concepts to better handle frequently asked customer questions. The research approach is described by defining objectives and research questions, followed by the used research methods. At last, a thesis outline is provided.

## 1.1    Problem statement

A QAS is a system that attempts to answer questions posed by users in natural language (Mishra & Jain, 2016). Information from various data sources is retrieved and processed to automatically provide a user with a single answer that is supposed to satisfy the user's need. In recent times, research towards QAS has received attention for the purposes of information retrieval (IR) and information extraction (IE) (Dumais, Banko, Brill, Lin, & Ng, 2002). A QAS combines IR and IE technologies to provide specific and brief answers (Athira, Sreeja, & Reghuraj, 2013). A wide variety of linguistic resources (like part-of-speech (POS) tagging, syntactic parsing, semantic relations, named entity extraction, dictionaries) have enabled various backbones for QASs, but the goal of a QAS has remained the same: retrieve answers to questions rather than full documents or best matching passages like an IR system does (Brill, Dumais, & Banko, 2002). Research on QASs is therefore mainly focused on achieving the highest accuracy in answering questions by improving existing techniques or introducing new techniques to enhance question processing, document processing, and answer processing.

### Research perspective
While the core functionality of QASs can be utilized to automatically answer frequently asked customer questions (high-volume, generally low-intelligent), research on QASs purposed for such customer service purposes is scarce. This solution has a two-fold benefit: (1) customers have a solution to self-service them that is more effective and efficient compared to traditional site search, and (2) answering a portion of customer questions automatically would relieve the customer service representatives (CSRs), which frees their time to manually solve the more complex customer questions. When deploying a QAS to handle frequently asked customer questions, users of the QAS are customers who are presumed not to be domain experts nor IT experts. Especially in that case, the experience of its usage to retrieve answers should be given attention in the system's design rather than its technical configuration only. Experience can be summarized as an individual's stream of perceptions, interpretations of those perceptions, and resulting emotions during one or multiple events (Roto, Law, Vermeeren, & Hoonhout, 2011). More specifically, UX can be summarized as a person's perceptions and responses that result from the use and/or anticipated use of a product, system, or service" (ISO 9241-210:2010, 2010). The UX field of study proposes means to design systems that enable positive UXs. However, in research on QASs, attention for the experience of using such systems is lacking. Therefore, this research is motivated by the lack of insights on how to enrich a QAS with UX concepts. In the context of this research, this QAS is purposed to handle frequently asked customer questions. It is hypothesized that enriching a QAS with UX concepts results in an improved UX, and therefore a better handling of frequently asked customer questions.

### Practical perspective
This master thesis research is conducted at ExperienceBank (from now on shortened as XBank), a fictive name that is chosen for confidentiality reasons. XBank is a Dutch international financial services provider, offering retail banking, wholesale banking, private banking, leasing, and real estate services. XBank serves more than 8 million customers around the world of which more than 7 million Dutch customers. XBank wants to ease their customer service department by developing a virtual assistant that can be of help for their customer services. This, to lower the workload and costs of the customer service department by lowering the number of CSR-customer interactions (by chat, phone, or contact form). Ultimately, the virtual assistant is smart enough to advice customers single-handed, and customer service processes get integrated so that CSRs get redundant. However, XBank is still in its infancy when it comes to a virtual assistant that interacts with customers. Currently, XBank only offers a keyword-based search functionality on their website. Therefore, the types of tasks that the virtual

assistant should be able to support are primarily the more simplistic tasks, like the request for a new debit card or providing information about costs for payments abroad (high-volume and low-intelligent requests or questions). To a certain extent, these types of tasks can already be fulfilled without interpersonal contact with a CSR by using the XBank website or the mobile app, but customers might still have trouble in finding the right webpage to fulfill their tasks. XBank has therefore the desire to offer an accessible virtual assistant that can answer the more simplistic questions instantly without interference of a CSR. XBank believes that a QAS is fitting for this purpose because of its capability to provide an automated response to a (simplistic) question of the customer within seconds, at any time. It is seen as an improvement compared to the current keyword-based search functionality. A QAS should reduce the number of interactions between CSRs and customers. However, the experience of a customer is of importance. XBank has set a standard for the customer experience the company should propagate. XBank strives for a Customer Satisfaction score of above 9 which they call the '9+ customer experience'. It means that every employee should deal empathically with customers, provide them a personalized service, and surprise and amaze them with the service they deliver. These core drivers that enable the '9+ customer experience' may be hard to adhere to with a QAS since interpersonal interaction is lacking. 'Personal contact', 'involvement', 'human-centered communication', 'personal support during hard times', and 'thinking with the customer' are examples of values that XBank strives for. A QAS, with its basic input-output mechanism, seems unable to fulfill these foremost human characteristics. Therefore, it is of interest of XBank to research the possibilities for enriching a QAS with UX concepts to deliver the same or a better experience to the customer when providing information through question-answering automatically. This means that it must be clarified which elements can enhance the experience and how to address experience elements in a QAS.

## 1.2 Research approach

In this section, the main research objective is formulated, the accompanying research scope is defined, and the research questions are presented and described. Next, the research method that guides the research is described: the design cycle of the Design Science Method (Wieringa, 2014). An explanation of the activities per phase is given. Thereafter, the used research methods are described. Lastly, the thesis outline is provided.

### 1.2.1 Research objectives

The main objective is to research how a QAS can be enriched with UX concepts to better handle frequently asked customer questions. To clarify this objective, the scope of this research is defined:
- A QAS is to be interpreted as a mechanism that takes a single question into consideration to produce a single answer as a response. This is different from question-answering through a chatbot, in which a dialog is produced. Dialogs are not in scope.
- The use of the word 'handling' is intentionally because customer questions posed to the QAS are either to be answered or be redirected (i.e. a way out) since answering questions automatically with a QAS is dependent of the available content to retrieve answers from.
- In this research, the focus is on frequently asked customer questions (i.e. FAQ), which are high-volume and generally low-intelligent questions.

Furthermore:
- The context of the QAS in this research is a customer service context of a Dutch commercial organization. The intended end users are therefore customers, who are presumed not to be domain experts nor IT experts.
- This research is not intended to realize a QAS and test, compare, or improve techniques. The focus is on the enrichments for a QAS. Therefore, this research is not intended to result in a (design of a) functional QAS. It is intended to elaborate on possible enrichments for a QAS from a UX perspective.
- Theory on UX provides means to design a system that enables a particular UX and will therefore be consulted for concepts that can enrich the QAS in its handling of frequently asked customer questions. In this research, concepts from related fields of study that also acknowledge the experience of customers are out of scope. The research is purely focused on enriching a QAS with concepts directly from the UX field of study.

Following the main objective, the main research question is defined as follows:

*How can a question-answering system be enriched with user experience concepts*
*to better handle frequently asked customer questions?*

To guide this research and fulfill its objective, research questions are formulated, and explanations of their purpose are given in Table 1.

*Table 1. Research questions.*

| RQ1 | *What is a question-answering system and how does it support handling frequently asked customer questions?* |
|---|---|
| | Answering this question should clarify what QASs are, what they are capable of, and what types of configurations exist. This requires an understanding of characteristics and methods of QASs to consider for a QAS purposed for handling frequently asked customer questions. |
| RQ2 | *What is user experience, what are its concepts, and how does it relate to handling frequently asked customer questions?* |
| | Answering this question should clarify what UX is, what UX concepts exist, what their goals are, and how they relate to handling frequently asked customer questions. This requires an understanding of existing UX concepts to argue what UX concepts can enrich a QAS and should therefore be considered. |
| RQ3 | *How can user experience concepts be covered in a question-answering system?* |
| | Answering this question should clarify how UX concepts can enrich a QAS and how they should be intertwined in a QAS. This results in descriptions of how UX should be conveyed in a QAS purposed for handling frequently asked customer questions which can be validated. |

## 1.2.2 Design Science Method

To conduct this research, the Design Science Method by Wieringa (2014) is used. It guides the investigation of the problem situation, the design of a treatment for the problem situation, and the validation of the treatment for the problem situation. This is known as the design cycle. The required actions are presented per phase in Table 2.

*Table 2. Design cycle phases.*

| Problem Investigation | Treatment Design | Treatment Validation |
|---|---|---|
| **PI1**: Identify stakeholders. | **TD1**: Specify requirements and context assumptions for the treatment to enrich a QAS for handling frequently asked customer questions with UX concepts (based on PI3 and PI4). | **TV1**: Validate the treatment's capability to enrich a QAS for handling frequently asked customer questions with UX concepts (TD3). |
| **PI2**: Identify stakeholder goals for question-answering and its experience (PI1). | **TD2**: Validate the contribution of specified requirements and context assumptions (TD1) to the stakeholder goals (PI2). | **TV2**: Validate the requirements of the treatment (comparing TD1 with TV1). |
| **PI3**: Define a QAS and explore existing QAS characteristics and methods. | **TD3**: Design the treatment for enriching a QAS for handling frequently asked customer questions with UX concepts (based on TD1). | |
| **PI4**: Define UX, UX concepts, and explore its relation to handling frequently asked customer questions. | **TD4**: Develop a PoC of a QAS for handling frequently asked customer questions to exemplify the application of the treatment (TD3). | |

### *Problem investigation*

The problem investigation phase requires a literature review and an empirical investigation for the identification of stakeholders and their goals in the context of question-answering as a customer service and its experience. Literature on existing QASs, QAS characteristics and methods, and related solutions results in insights for the design of a QAS that can handle frequently asked customer questions. This is complimented with empirical

insights based on market research towards Dutch online QASs. For an understanding of the body of knowledge on UX and its relation to handling frequently asked customer questions, literature in this domain must be reviewed. This is complimented with empirical insights based on knowledge of XBank employees regarding the experience of users in the context of automated question-answering. Part A of this thesis covers the problem investigation phase.

### *Treatment design*
The treatment design phase builds upon findings from the problem investigation. These findings enable the creative process of designing the treatment that facilitates UX enrichment in a QAS. Next to this treatment, a QAS PoC is developed to exemplify its application. Part B of this thesis covers the treatment design phase.

### *Treatment validation*
The treatment validation phase requires a validation by expert opinions to determine whether the treatment leads to successful UX enrichment in a QAS and an evaluation of the QAS PoC demonstrating the treatment application. Whether the treatment meets the initial requirements must be validated as well. Part C of this thesis covers the treatment validation phase.

## 1.2.3 Research methods

The used research methods are described in this paragraph, covering a literature review, an empirical investigation, and expert opinions.

### *Literature review*
To perform a literature review, the systematic literature review as described by Kitchenham (2004) is used. A systematic review is the identification, evaluation, and interpretation of all available research relevant to a research question (or topic area or phenomenon of interest). She states that the most common reasons for undertaking systematic reviews are to summarize existing evidence concerning a treatment or technology, to identify gaps in current research (which suggests areas for further investigation), and to provide a framework or background to appropriately position new research activities.

The search for literature is performed with search strings (consisting of relevant keywords) on various literature databases. Reviewing the literature will result in clarity about the research gap that justifies this research. Definitions in the areas of QASs and UX are to be clarified, and insights about QASs (e.g. what are the different types, available technologies, architectures) and UX (e.g. what is understood by UX, aspects and factors of influence) are gained by reviewing literature from the past and looking at the current state of research in these areas. These insights are required to make informed decisions for the enrichment of a QAS purposed for handling frequently asked customer questions.

Various keywords are used to find literature about the main concepts. The keywords to search for literature in online literature databases are presented in Table 3 and in Table 4. The databases (or search engines) that are consulted are ACM Digital Library, Google Scholar, IEEE Computer Society Digital Library, and ScienceDirect.

*Table 3. Keywords regarding question-answering systems.*

| Main keywords | | Main keyword(s) + | Main keyword(s) + system + |
|---|---|---|---|
| Question | Answer | System, Application, Software | Architecture, Design |
| | Answering | Automated, Automatic, Automation | Technology, Techniques |
| | | Survey, Review | Information retrieval, Information extraction |
| | | Approaches | Natural language processing |
| | | FAQ, Frequently asked questions | Ontology |
| | | Dutch | Restricted domain, Closed domain |
| | | Human computer interaction | Survey, Review |
| | | | Approaches |
| | | | FAQ, Frequently asked questions |
| | | | Dutch |
| | | | Human computer interaction |

*Table 4. Keywords regarding user experience.*

| Main keywords | | Main keyword(s) + |
|---|---|---|
| User | Experience | Online, Digital |
| Customer | Satisfaction | Quality |
| | Service | Performance |
| | Support | Perception, Perceived |
| | | Model, Framework |
| | | Metrics, Parameters |
| | | Questionnaire |
| | | Survey, Review |
| | | Design |
| | | Approaches |
| | | Self-service |
| | | Helpdesk, Servicedesk |
| | | Human computer interaction |

Single keywords are used but also combination, like suggested in both tables. Combinations of QAS-related keywords and UX-related keywords are required to search for crossover literature (e.g. "question answering system" combined with "user experience"). For the search strings, the AND- and OR-operators should be used, and quotes are used for searching for exact terms or paraphrases. Filtering options are used, for example, to focus the search for recent research only.

In addition to the search method for the systematic literature review, the snowballing technique is also used to gather knowledge from literature. This approach aims to work from a start set of papers and perform iterations of backward and forward snowballing (Wohlin, 2014). Backward snowballing means that the reference list of a paper is used to identify new papers to review, based on a selection criterion. Forward snowballing follows the same procedure but focuses on papers that cited the paper at hand. Snowballing overcomes the problem of not discovering relevant and important papers, that may be of high-quality. This, because these papers may be hard to identify by performing database searches only. Papers that are found and selected for the initial review are used as start papers to discover more literature through backward and forward snowballing.

### *Empirical investigation*
The empirical investigation is required to gather insights from practice. Unstructured interviews are conducted with employees of XBank who are familiar with the customer service domain, its digital solutions, and the accompanying UX. This, to gather information regarding question-answering (via various channels) within XBank and developments for its automation, and views on UX in the context of (automated) question-answering. Furthermore, market research is conducted for an understanding of current online QASs of Dutch commercial organizations to define a common practice Dutch online QAS.

### *Expert opinions*
The designed treatment is validated by expert opinions as described by Wieringa (2014) to gather qualitative data. Validation by expert opinions means that the guidelines are presented to multiple experts who imagine how the treatment would work in practice. This leads to predictions of the effects and thus points of adequacy and shortcomings for the designed treatment. The experts' imaginations of the artifact in the problem context and their elaboration on the predicted effects serves as the validation model which is observed. Because the imagination of experts is of importance, the treatment and foremost its application should be explained in a realistic problem context. If this is not the case, mainly positive opinions are often a result because it is hard for them to imagine its practical application. To prevent this, experts are asked to elaborate their opinions. Indications for improvement opportunities for the treatment are logically foremost a result of negative opinions and those are therefore more useful than positive opinions. It will point out the bad design ideas.

The goal is to have at least five experts, classified as experts on UX based on their function and years of employment at XBank. Conducting the validation with five experts leads to a broader picture due to multiple perspectives. The experts are required to have knowledge of UX and familiarity with XBank's customer service domain, to expect that they possess knowledge on factors that affect a certain experience of the customer when using a system. The experts are approached within XBank since multiple employees fulfill these types of roles.

### 1.2.4 Thesis outline

The thesis is divided into three parts based on the Design Science Method phases (Wieringa, 2014): part A: problem investigation, part B: treatment design, and part C: treatment validation. In part A, the conducted literature review (chapter 2) and empirical investigation (chapter 3) are presented. Thereafter, all the findings from the former chapters are recapitulated to prospect on the to be designed treatment (chapter 4). In part B, the designed *user experience guidelines for question-answering systems* are presented (chapter 5). In part C, the conducted validation is described first (chapter 6), followed by a presentation of its results (chapter 7). After these chapters, the discussion (chapter 8) and conclusion (chapter 9) are presented. Thereafter, only the bibliography and appendices are to be found.

### 1.2.4 Thesis outline

# PART A: PROBLEM INVESTIGATION

# 2    LITERATURE REVIEW

In this chapter, the theory that has been consulted by reviewing the literature related to the main subjects of this research, QASs and UX, is discussed. This chapter will first elaborate on QASs followed by an elaboration of UX. Finally, the main findings and conclusions are drawn.

## 2.1   Question-answering systems

This section discusses QASs. First, QASs are introduced and explained, followed by a description of its general architecture. Then, several examples of QASs applied in practice are shortly described. An extensive paragraph follows, covering classifications of QASs based on the work of Mishra & Jain (2016). These classifications are discussed to understand QAS configurations. Next, techniques that were not covered in the classification but are found relevant are discussed. After that, the possibilities for Dutch natural language processing (NLP) are reviewed. Thereafter, evaluation metrics for produced answers are listed as means to measure the quality of produced answers. Finally, the existing research on the experience side of QASs is discussed. The section ends with a summary of the discussed content.

QASs attempt to answer questions posed by users in natural language (Mishra & Jain, 2016). Information from various data sources is retrieved and processed to automatically provide a user with a single answer that is supposed to satisfy the user's need. Research towards natural language question answering by computer finds its origins in 1959, and the first QAS with underlying NLP techniques were developed in the following years (Simmons, 1965). These systems attempted to answer textual English questions. In more recent times, research towards QAS has received attention for the purposes of IR and IE (Dumais, Banko, Brill, Lin, & Ng, 2002). An IR system presents a list of documents that relate to the user's information need, and an IE system extracts explicitly stated information of interest from one or more texts provided in the domain (which is well defined), and structures and combines the found data. A QAS combines these technologies to provide specific and brief answers (Athira, Sreeja, & Reghuraj, 2013). A wide variety of linguistic resources (like part-of-speech (POS) tagging, syntactic parsing, semantic relations, named entity extraction, dictionaries) have enabled various backbones for QASs, but the goal of a QAS has remained the same: retrieve answers to questions rather than full documents or best matching passages like an IR system does (Brill, Dumais, & Banko, 2002). The challenge is being accurate with provided answers and preventing incorrect answers. Therefore, the QAS is mainly to be designed for the domain it is purposed for, but there are more aspects that define a suiting design, e.g. the type of user or the type of question.

### 2.1.1  Question-answering system architecture

Architectures of researched QASs vary. These variations (e.g. additional or different modules) are caused by the possible variation in characteristics (e.g. its purpose domain, the approach, applied techniques and methods). The most prominent reason for adjustments in the architecture is the improvement of the accuracy of the system. Question-answering is always the central goal of a QAS, and changes in parts of the architecture might improve the process to achieve that goal. Looking at research towards QASs, ranging from 2000 until now, it can be concluded that the general architecture is a fairly standard pipeline of three modules (also called components), each purposed for an accompanying phase of the whole question-answering process. These modules, in general, are required for phases of (1) *question processing* (or *question analysis*), (2), *document processing* (or *document analysis*), and (3) *answer processing* (or *answer analysis*) (Hovy, Gerber, Hermjakob, Junk, & Lin, 2000; Jijkoun, Mishne, & de Rijke, 2003; Wang, 2006; Kolomiyets & Moens, 2011; Allam & Haggag, 2012; Gupta & Gupta, 2012; Athira, Sreeja, & Reghuraj, 2013; Dwivedi & Singh, 2013; Yogish, Manjunath, & Hegadi, 2016). A visualization of the general architecture is shown in Figure 1.
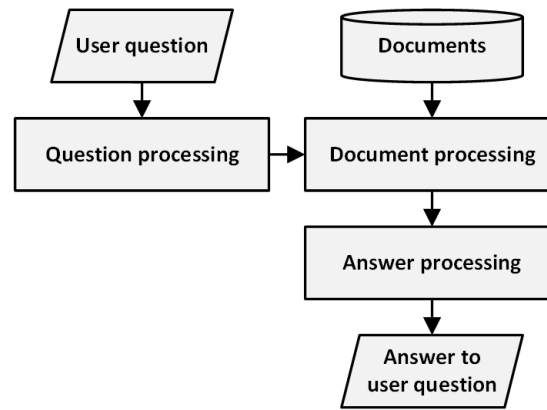
*Figure 1. General architecture of a question-answering system.*

### Question processing

The module for *question processing* is concerned with analyzing, classifying, and reformulating the question of the user. Input for this module is the question from the user, textually and in natural language. The question must be dissected to make it understandable for the system. Transforming the question content into a query form is a necessity for retrieving matching information from a data source. Also, defining the type of question can be helpful in retrieving fitting information. Therefore, the output of this module usually is a system-readable question and the question type.

### Document processing

The module for *document processing* is concerned with retrieving, filtering, and ordering documents in which answers must be sought. The input is the system-readable question and its type from the first phase, and these are used to search for relevant information that might answer and suit the question. This module requires IR techniques to fulfill that goal of finding information. Documents must have a certain representation model to have the system perform searches with a dedicated technique. This is similar to the function translating a question into a certain query language format: the system must be able to make sense of the content. Documents that may contain the answer are extracted and form the output of the document processing phase.

### Answer processing

The module for *answer processing* is concerned with identifying, extracting, and validating the answer to the question. The documents that were found by searching for relevant information along with the question are the input. From the documents, answer candidates are extracted based on the requested information in the question. Again, IR techniques are required, but this time to fulfill the goal of finding the candidate answers. From those candidate answers, a final answer must be chosen. This requires ranking of candidate answers. The answer candidate with the highest probability of answering the user question is to be presented to the user. Most often, this is a literal passage or entity from the document. This last module is the most essential part of a QAS because this overcomes having to present (a) full document(s) with possibly the information that the user is looking for. Instead, only an extraction from a document, that has a high probability of answering the initial question, is presented. This is the main difference between an IR system and a QAS.

## 2.1.2 Examples of question-answering systems in practice

In this paragraph, examples of QASs applied in practice that are found in literature are briefly described to provide context about the application of these systems.

### Helpdesk context

Samarakoon, Kumarawadu, & Pulasinghe (2011) developed a closed domain QAS to automate a customer helpdesk of a commercial organization. This knowledge-based QAS is aimed at maximally using the experience with users available with CSRs of the organization rather than application data. This approach considers syntactic, lexical, and morphological variations and performs synonym transduction. The query understanding method is based on a ranking algorithm and pattern writing process. The intent, context, and content components of natural language and the word order are considered in this pattern writing process. The pattern writer requires expertise in UX, to understand the language characteristics of users in a specific application domain. Collections

of (transcribed voice) dialogs between users and CSRs are used to write patterns, which are to be verified by another pattern writer to minimize inconsistency between patterns. The system uses a threshold to decide upon answering these new questions, because it is assumed that with customer service, it is worse to provide a wrong answer than no answer at all.

### *Clinical context*

Cao, et al. (2011) developed a clinical QAS for robust semantic analysis on complex clinical questions. As answers, the system delivers question-focused extractive summaries. Innovative approaches for question analysis (supervised machine-learning to classify questions by topics and identifying keywords), summarization (clustering technique with extraction of content-rich keywords for hierarchical structured summarized answers), and answer presentation (hierarchical answer presentation interface with grouped passages per topic based on two-layer clustering) are implemented. The system allowed physicians to enter questions in natural language with minimal query formulation and the results showed that information needs were quickly met due to efficiently navigation among answer sentences. The system has the potential to assist physicians in their work and thereby improve the quality of patient care.

### *Physics domain context*

Abdi, Idris, & Ahmad (2018) developed an English ontology-based QAS for the physics domain (Electricity and Electromagnetism). It enables the user to use natural language to retrieve information from formal ontologies through queries. The question processing component uses tokenization, stemming, part-of-speech tagging, stop word removal, annotations, and ontology lexicon. For answering, the QAS uses an inferring schema mapping as its backbone, which acts as an intermediate language for mapping natural language queries to an ontology-compliant query. This inferring schema mapping combines semantic and syntactic information and attribute-based inference to do so. For building the ontology, relevant standards and regulations were utilized extensively. The QAS was built in such a way that it can be applied in other domains and languages as well by using an ontology model and a query database of a different domain. The QAS's performance level proved to be useable in real environments.

## 2.1.3 Classifications of question-answering systems

Mishra & Jain (2016) have surveyed literature on QASs and created a QAS classification. This resulted in eight ways to classify a QAS. These classifications are:
1.  Application domains for which QASs are developed
2.  Types of questions asked by a user
3.  Types of analyses performed on questions and documents
4.  Types of data consulted in data sources
5.  Characteristics of data sources
6.  Types of matching functions of different retrieval models
7.  Forms of answers
8.  Types of techniques used for answer retrieval.

This classification is used as a guideline for a more in-depth discussion of QASs to get an overview of existing variations. Before discussing the classifications, Figure 2 shows how each classification type (dashed boxes) in general relates to the general architecture of a QAS. It shows that the application domain is decisive for everything involved with a QAS, and that the technique for answer retrieval concerns all QAS modules, while question type, question and document analysis type, data source type, data source characteristics, matching function type, and answer form only concern specific QAS components. Colors are used for better visibility.
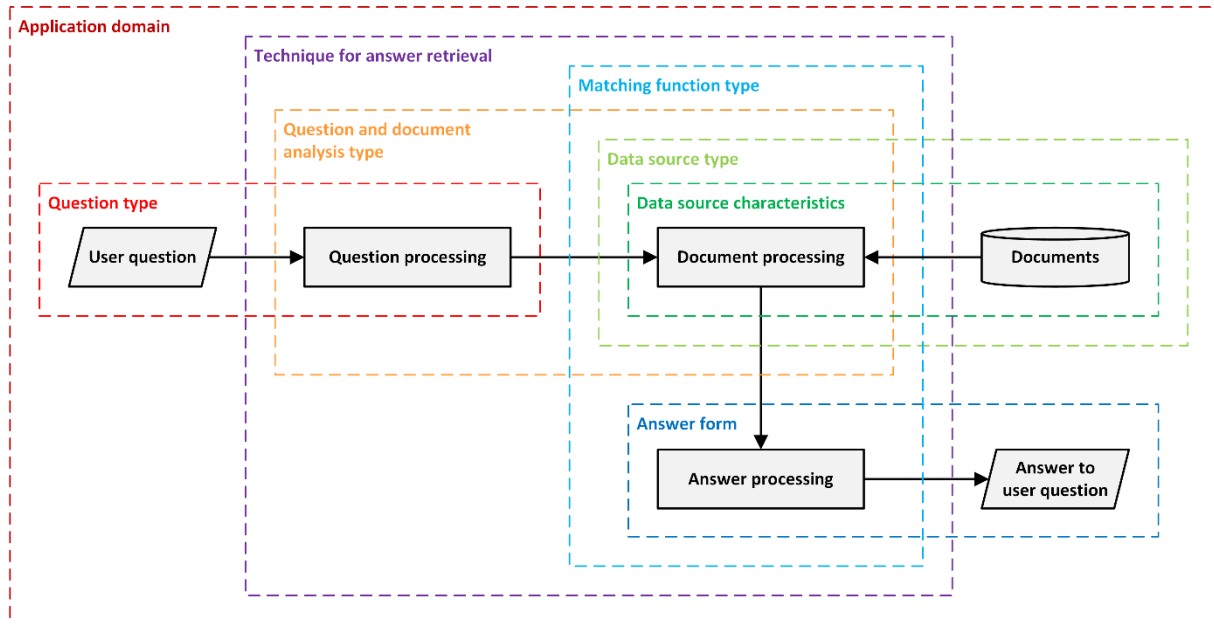
*Figure 2. Classifications of question-answering systems in relation to the general architecture.*

## Application domain

The most distinctive classification of a QAS is its application domain. Two main types can be identified: the *open domain* (sometimes referred to as *general* domain) and the *restricted domain* (sometimes referred to as *closed domain*).

1. An *open domain* QAS is domain independent and exploits a general ontology and world knowledge. This means that the QAS strives to answer a large repository of question asked by its users and uses as much available information sources as possible to give an answer in response. The web often provides the backbone for an open domain QAS because of its tremendous amount of available data (Dumais, Banko, Brill, Lin, & Ng, 2002). An example is the use of all Wikipedia articles (Chen, Fisch, Weston, & Bordes, 2017) or news articles. This kind of QAS is usually not able to offer high-quality answers and is meant for casual users. It is characterized by its lack of a domain specific dictionary and therefore, users are not required to have knowledge of domain specific keywords. The satisfaction of using an open domain QAS is user dependent because goals and expectations can differ among users. For instance, domain experts will often require specialized information while an open domain QAS strives to deliver simple answers for diverse questions instead of detailed, domain specific answers.

2. The *restricted domain,* however, is specifically meant for domain specific questions, and answers to those questions are sought after in domain specific document collections with specialized extraction rules (Diekema, Yilmazel, & Liddy, 2004). This limits the repository to find question patterns and therefore answering is limited, but it does offer high accuracy in answering and higher quality answers. Domain specific ontology and terminology are required. Its relative smaller size makes it possible to apply more complex NLP techniques offline (Mollá & Vicedo, 2007). This suits domain experts, who are assisted with specialized answers. Satisfaction is still user dependent, but a restricted domain QAS strives to assist users from a specific domain only which lead to a better match between users and its answering capabilities.

## Question type

The types of questions a QAS strives to answer is highly related to the task of generating answers because that process is directly affected by question classification. Mishra & Jain (2016) name six types of questions: *factoid type*, *list type*, *hypothetical type*, *causal*, *confirmation*, and *opinion*.

1. *Factoid type* questions are what-, when-, which-, who-, and how-questions, typified as simple and fact-based and generally used for named entity questions. Entity tagging software achieves a good accuracy in tracing entities and currently offers a satisfactory performance. A large repository of factoid questions can be asked and there is no need for complex NLP processing. The expected answers are short phrases or sentences. It requires a fact expressed in the text body as an answer (Kolomiyets & Moens, 2011).

2. *List type* questions are like the factoid type, but the answer is a list of entities or facts (Kolomiyets & Moens, 2011) instead of one. It is seen as a series of factoid questions, but with each new question the previous given answer is ignored. Similar techniques can be applied, and NLP is again not necessarily required for extracting answers. A problem can be that the questions asked may not withhold (a) clear threshold value(s).

3. *Hypothetical type* questions request information about hypothetical events. This results in subjective answers and therefore no specific correct answers. It can be used for exploring various answers based on world knowledge and common-sense reasoning. The trade-off is that answers are diffused, and the accuracy and reliability are low.

4. *Causal* questions are how- or why-questions. It is used for explanations, reasons, or elaborations about specific events or objects. This requires advanced NLP techniques to analyze text at a pragmatic and discourse level. The challenge lies in determining relevant or unique answers to subjective questions. Those answers range from a sentence to a paragraph and up to a whole document. It is important to accurately extract passages or phrases as answers that bear causes (Higashinaka & Isozaki, 2008). It is challenging to base the answer on the intention of the user meaning that the user question needs to be interpreted correctly. Retrieval models based on the bag-of-words model is faced with problems due to polysemy, homonymy, and synonymy. For answer generation, it is required to identify the discourse relationships in source documents.

5. *Confirmation* questions require "yes" or "no" answers. This requires an inference mechanism, world knowledge, and common-sense reasoning. However, a higher level of knowledge acquisition is required and retrieval techniques which are still in development.

6. *Opinion* questions target subjective information about an entity or event. The social web (Web 2.0) can be used in combination with opinion mining techniques, which is the process of extracting opinions from text documents (Missen, Boughanem, & Cabanac, 2013). The massively generated user data on the web (e.g. networking sites, blogs, reviews) is opiniated data with public opinions, which can assist users in judging products. However, questions are often informal, and it is difficult to process them due to poor semantics. It is also challenging to classify text as objective or subjective and to find relevant opiniated documents. These texts are often difficult to process due to lack of grammatical punctuation. Also, a problem is the existence of fake or spam content because it affects the mined opinion.

### Question and document analysis type

Various techniques exist to analyze the question from a user and analyze the content of a document. Mishra & Jain (2016) define six analysis types: *morphological*, *syntactical*, *semantic*, *pragmatic and discourse*, *expected answer*, and *focus recognition*. Currently, some of them are only applicable for question analysis.

1. *Morphological* analysis is a process of separating words into individual morphemes and assigning a class to the morpheme. In linguistics, morphology refers to the internal structure of words and the systematic form-meaning correspondences between words (Booij, 2012). This may give useful clues as to the word's category (Bird, Klein, & Loper, 2009). Stemming and lemmatization of words are required to do a morphological text analysis. This analysis type considers different forms of words which removes redundancy at word level during the IR process. This leads to effective search. However, stemming could yield incorrect results because of words with the same stem word but a different semantic.

2. *Syntactic* analysis is the identification of the grammatical construction of words in a question and in documents. Syntax is defined as the set of rules or principles that govern how words are put together to form phrases, which are well formed sequences of words (Sportiche, Koopman, & Stabler, 2013). Syntactic analysis is done because sentences usually contain content bearing keywords (nouns, verbs, adjectives, adverbs) and function words. The process results in parse trees. It is an attempt to reduce the search space in documents which enhances the search effectiveness. This analysis type considers words of different parts of speech which removes redundancy at word level during the IR process. A challenge is the syntactic ambiguity in questions and the possibility of incorrect interpretations of questions.

3. *Semantic* analysis is the deduction of possible meanings of a question based on the words used in the question. Semantic analyses demonstrated that tools of linear algebra can extract automatically what is conveyed in text, instead of having to translate texts to a formal notation (Landauer, McNamara, Dennis, & Kintsch, 2007). This helps finding the semantic class of question (and answer) types. It is generally a follow up of the syntactical analysis, because it analyzes the generated parse trees and then interprets the possible meaning of the question. This provides a more effective search compared to keyword-based search. An additional task is semantic role labeling, which tries to identify arguments in a text. According to Mishra & Jain (2016), semantic analysis operates at lexical and sentence level only for deducing the

meaning (not on document level), and the problems that occur with co-reference resolution, named entity recognition, relation extraction, and part-of-speech (POS) tagging make it difficult to do semantic analysis of text. However, deep learning techniques have proven to consider semantics of words, phrases, paragraphs, and even documents very well compared to traditional techniques (Menger, Scheepers, & Spruit, 2018). The deep learning technique that uses question and answer embeddings is elaborated in paragraph 2.1.4. Also, research on performing semantic role labeling without syntactic knowledge through the use of recurrent neural networks has proven to outperform previous combinations with different parsing trees or models (Zhou & Xu, 2015).

4. *Pragmatic and discourse* analysis is the interpretation of questions and documents at sentence level or higher. It can be seen as a broader syntactical analysis, since it is the pragmatic interpretation of utterance and context with which the sentence is expressed instead of only the sentence itself. Discourse analysis identifies the discourse structure of the connected text to retrieve the type of discourse relationship that exists between sentences. This analysis type is preferred for long and complex questions which require long answers, because it can help tracing answers by finding relations among sentences (e.g. elaboration, explanation, contrast). Anaphora resolution (replacement of words which are semantically empty) and discourse structure recognition (identification of logical connectivity of sentences) are tasks of discourse analysis. The latter is specifically required for opiniated, causal, hypothetical, and confirmation type questions. However, technology that uses (intra) sentential discourse structure of sentences is still in its infancy, and the task is hard because of problems in co reference resolution, named entity recognition, relation extraction, and part-of-speech tagging.

5. *Expected answer* analysis is the determination of the entity that is required in the answer, based on the question type. This mainly helps for factoid and list type questions.

6. *Focus recognition* analysis is the recognition of focus in the question. This important aspect can have a significant effect in finding correct answers.

In addition to the described question and document analysis types, Mishra & Jain (2016) also state the distinction of three approaches for the analysis of questions and source documents: *statistical based*, *rule-based pattern matching*, and *hybrid*. In addition, Dwivedi & Singh (2013) also define a *linguistic* approach.

1. The *statistical based* approach is a data driven approach that uses quantitative relations to discover statistical relations existing in questions and documents. This includes probabilistic modeling, linear algebra, and information theory. Statistical queries mostly consist of keywords from the question instead of artificial query languages. It does not require expert knowledge, natural language problems (e.g. grammar, paraphrasing) are ignored, can easily handle large amounts of data, can deal with complex questions, and it can deal with heterogeneous data sources. This statistical approach offers promising results, but it does require a large amount of data for correct statistical learning. It does disregard any semantics or context of words and sentences. However, deep learning techniques, which are statistical based, have proven to address semantics through question and answer embeddings (see paragraph 2.1.4).

2. The *rule-based pattern matching* approaches make a combined use of linguistic rules and human knowledge in IR processes. It works with predefined patterns (lexico-syntactic or lexico-semantic) built for questions and answers. The extraction of documents with answers is based on matching questions to those patterns. This approach requires less data compared to the statistical approach, but it can also easily handle large amounts of data and deal with complex questions. It does, however, require expert (or domain) knowledge. Building the patterns is cumbersome and non-trivial, and it can be hard to create the correct patterns because natural language does not always follow a definite pattern.

3. The *hybrid* approach combines the two. Rule-based pattern matching approaches require bootstrapping or initial clustering using statistical methods, and statistical based approaches require large amounts of data for correct learning. The hybrid approaches minimize the observed limitations, but requires more data compared to rule-based pattern matching approaches.

4. Dwivedi & Singh (2013) define a *linguistic* approach, which relies on artificial intelligence-based methods that integrate NLP techniques and a knowledge base or corpus, to build question-answering logic. Linguistic techniques (tokenization, part-of-speech tagging, parsing) are used to formulate user questions into precise queries to extract answers from structured databases. This approach is generally applied to systems for long-term information needs for a particular domain only since it has the requirement of a domain specific knowledge base. Portability of the QAS is limited and building an appropriate knowledge base is time-consuming.

### Data source type

Data source types are decisive for the possibilities of answer retrieval in (a) data source(s). Mishra & Jain (2016) name three types of data sources: *structured*, *semi-structured*, and *unstructured*.

1. A *structured* data source has data structured in a semantic set. Similar entities are collected in relations and entities in the same relation have the same attributes. The data is arranged in a defined format. Therefore, the matching between a query and the data is exact, but the required query language to do so is artificial. The reliability of answers is high, given the fact that it comes from information stored in a specific data source and data structure, and such a data source does not require complex NLP. However, storing information is limited, reference reconciliation can occur, building it is labor intensive, different structured data sources exist (e.g. MySQL, SQLite, DB2) with different representations and query languages, and natural language questions must be transformed into the dedicated query type.

2. A *semi-structured* data source has no partition in between stored data and the schema. Representation of information is constrained by the schema, provides a flexible format for data exchanges in between different query languages, and transforms structured data into a semi-structured format. Again, reference reconciliation can occur and building a semi-structured data source is also labor intensive to build.

3. An *unstructured* data source contains data of any type. It is not structured in any semantic set and there are no strict rules for the arrangement of data. The use of NLP and IR techniques are necessary when dealing with unstructured documents to find answers. Information can easily be added or updated. However, representation is a problem, reliability of answers is low, and paraphrasing is prominent.

### Data source characteristics

Data sources are typified by their characteristics, which influence the data processing possibilities. Mishra & Jain (2016) name five: *source size*, *language*, *heterogeneity*, *genre*, and *media*.

1. The *source size* affects the range and richness of information and the accuracy of finding answers. Events and objects will have better descriptions when the size is larger, which leads to more satisfactory searching for answers. As stated before, a larger size will result in better accuracy for statistical and rule-based pattern matching approaches because of more data for learning.

2. *Language* is of influence when multilingual documents are used. It can provide more knowledge when multilingual sources are combined but it is a challenging task. Different languages follow different syntax rules and grammar meaning that there are no common linguistic rules to dissect multilingual documents. It requires different language processing techniques.

3. *Heterogeneity* is concerned with large amounts of information stored among different sites and in different formats. The information scattered in different formats (e.g. database, text, multimedia) can provide more knowledge if it is integrated but this is challenging. There is no common representation model to model different types of data sources. Different data sources often require different query languages, which, as stated before, is a problem for the transformation from a natural language into a query. It is also difficult to populate the data space with new (different) data sources based on their information content.

4. The *genre* of a data source is the distinction between the use of linguistically correct (formal) language and linguistically incorrect (informal) language. Informal language will result in (somewhat) incorrect parse trees, which makes it difficult to retrieve answers.

5. *Media* other than text, such as audio and video, are still difficult to use for answer retrieval.

### Matching function type

To retrieve answers, some type of matching function is required to match the content of the question with a fitting piece (or ranked pieces) of information (the answer) from the document collections. Mishra & Jain (2016) describe the matching functions of five different retrieval models: *set theoretic (standard Boolean)*, *algebraic*, *probability*, *feature-based*, and *conceptual graph-based*.

1. The *set theoretic (standard Boolean)* retrieval model treats documents as sets of words or phrases, and performs matching based on operations carried out in between those sets. It is an easy implementation because it is easy to implement keyword matching which uses the Boolean operator in making a query. However, partial matching of a query with a document is not done because systems are unable to rank the output if little matching occurred between question keywords and document keywords. Another restriction is the fact that a user must include Boolean expressions in the question instead of providing a natural language question. This can be difficult for users and cause aversion. Also, using the Boolean operator usually results in either too many or too few documents.

2. An *algebraic* retrieval model represents questions and source documents as vectors, matrices, or tuples. The matching is performed through scalar values in a vector space, making use of a generalized vector space model, a topic-based vector space model, latent semantic indexing, or an extended Boolean model. The most common used model is the bag-of-words model, but the deep learning technique that uses question and answer embeddings through vectors (see paragraph 2.1.4) seems a useful alternative.

3. A *probability* retrieval model uses probability relevance to treat questions and documents. This can be a binary independence model, a probabilistic relevance model, uncertain inference, or language models. This type can present a user a ranked list of documents after the matching process. Users can express the query in natural language instead of a technical query language. The expressiveness is however limited (e.g. Boolean operators such as 'NOT' cannot be used). There is also some lack in capabilities of representing some linguistic features (e.g. phrases or proximity constraints) for information search. Another point is that the required prior estimation of probabilities for words of questions and documents causes a costly relevancy computation.

4. A *feature-based* retrieval model determines a single relevance score based on features of documents (as a vector of values of feature functions). This enables the search for feature specific information. However, identifying values of features, functions, and their computation is a problem, which is required for learning to rank.

5. A *conceptual graph-based* retrieval model represents the sentence of a text with a structure of vertices and edges in a graph. This provides a higher-level understanding of the text because it captures its semantics. It does require questions and sentences to be modeled into a conceptual graph formalism to be able to retrieve answers with it, and it is difficult to model complex questions and documents and find the relevancy between them.

## Answer form

Answers can be presented to the user in different forms. Mishra & Jain (2016) define two main categories of answer forms: *extracted answer* and *generated answer*. Both categories have some variations.

1. An *extracted answer* is an answer that is literally extracted from a source document. It can be in the form of (a) sentence(s). Source documents are divided into individual sentences and the most relevant sentence is presented. This is generally used for factoid or confirmation questions. A common problem is the lack of boundaries in informal documents. The answer can also be in the form of a paragraph. Source documents are then divided into paragraphs and the most relevant paragraph is presented. A problem however is topicalization of the paragraph. The answer can also be presented in the form of multimedia which can be audio and video but as stated before, this is still difficult to realize.

2. A *generated answer* is an answer that is formed based on content from a source document. It can be in the form of "yes" or "no" after verification and justification. The answer can also be in the form of an opiniated answer or rating based on content about an object or its features. The answer can also be in the form of an answer to be presented through dialog.

## Technique for answer retrieval

Choosing a technique for answer retrieval is decisive for all discussed criteria, and therefore, this classification is more of an overall choice than one to combine with the others. Different techniques can be used for answer retrieval and each effectiveness depends on the scenario in which it is being used. The scenario depends on the complexity of the questions asked, the data sources, and the answer desires. Mishra & Jain (2016) have made a distinction between four types of techniques that are used in QASs: *data mining*, *IR*, *natural language understanding (NLU)*, and *knowledge retrieval and discovery*. These techniques are discussed on ten aspects to provide an overview (and comparison). The aspects are: *searching*, *matching*, *technology*, *form of answers*, *types of questions*, *relevancy*, *analysis technique*, *knowledge source*, *model used in its retrieval process*, and *reliability*. This is a generalized mapping of these aspects per answer retrieval technique.

1. A QAS based on *data mining* is meant for searching for factual data, with an exact matching. The used technologies are artificial intelligence and databases. Answers are short and objective because they are meant to answer simple-find questions. It performs pattern matching and syntactical analysis. The knowledge source is a database and the used retrieval model is bag-of-words. The reliability is considered to be very high since domain experts have designed the database schema.

2. A QAS based on *IR* is meant for searching through querying for factual information, with a best match matching goal. The used technologies are IR and NLP. Answers are mixed (both short and long, depending on the requested information) and subjective, trying to answer wh-questions (e.g. what, where, which, when). It performs relevancy ranking, pattern matching, and syntactical analysis. The knowledge source

is the syntactic web and the used retrieval model is bag-of-words. The reliability is considered to be low because the web contains lots of fake information.

3. A QAS based on *NLU* is meant for searching through querying for information that is either fact based or subjective opiniated, with a best match matching goal. The used technologies are NLP and NLU. Answers are mixed and subjective, trying to answer definitional questions. It performs relevancy ranking, pattern matching, syntactic analysis, and semantic analysis. The knowledge source is the syntactic and pragmatic web and the used retrieval model is bag-of-concepts. The reliability is considered to be low, again because of the existence of lots of fake information on the web.

4. A QAS based on *knowledge retrieval and discovery* is meant for understanding knowledge, creating knowledge, and searching for useful correct answers, with a best correct match matching goal. The used technologies are NLU, knowledge acquisition, and mining. Answers are mixed and subjective, trying to answer complex questions (e.g. how, why, hypothetical, confirmation). The knowledge source is the semantic and pragmatic web and the used retrieval model is bag-of-knowledge. The reliability is considered to be high.

## 2.1.4   Other question-answering techniques

In search for additional techniques (techniques that were not covered in the Mishra & Jain (2016) QAS survey paper) to handle (part of) the question-answering process, only two approaches have been found to be of significant importance because of its potential for the to be designed QAS.

### *FAQ answering through question similarity*
While the discussed techniques are focused on understanding natural language questions and documents to extract text from extensive document collections as answers, another approach might offer a solution to answer frequently asked customer questions. A QAS focused on answering frequently asked questions (FAQ) can utilize existing FAQ content (question-answer pairs). When a user asks a question, the question is compared to the existing questions available in the FAQ content, and if a similar question is found, the corresponding answer is returned to the user as the final answer (Song, Feng, Gu, & Wenyin, 2007). To calculate question similarity, a statistical similarity (dynamically formed vectors) and a semantical similarity (between words) can be determined. The combination proofed to achieve good performance.

Sneiders (2009) proposed an approach that uses a dedicated FAQ-database and question templates for each FAQ. Such a template holds linguistic signatures (i.e. language structures that must be, may be, or are not present) that resemble regular expressions and language structures of expected user queries. This, to make conclusions about the semantic distance between the query and question templates. The linguistic signature is a result of syntactic (phrases), morphological (word stemming), and lexical (synonym) analysis of expected queries. A limited reasoning is implemented to disregard low match-confidence FAQs if there are high match-confidence FAQs found as well.

A limitation of this approach is the lack of details being addressed accordingly. On the other hand, the answers and their format are controlled by the provider, and are as specific and complete as intended, and in the tone as desired. Another benefit of this approach is that new FAQ can be identified when questions proof to be unaddressed in the FAQ-database, to populate the database based on users' input, presumably satisfying future users. Also, answers can be of other media types than text as well.

### *Deep learning: question and answer embedding through vectors*
Minaee & Liu (2017) developed a deep learning-based model to answer questions automatically. Questions and answers are embedded by using neural probabilistic modeling, followed by training a deep similarity neural network to calculate similarity scores for question and answer pairs. The answer with the highest similarity score indicates the best answer to a question. Neural network-based techniques are used to find a word embedding that captures similarity among words, which is a limitation with, for example, the bag-of-words method because that ignores semantic information. The similarity among words is expressed as a relatively similar representation in the embedding space. In this research, the model is trained on a public question-answering database, followed by finetuning to use it on customer care chat data. Tests with the framework achieved very good performance.

Better performance in NLP tasks is achieved by grouping similar words. Mikolov, Sutskever, Chen, Corrado, & Dean (2013) introduced the Skip-gram model which efficiently learns high-quality vector representations of words (word vectors) from large amounts of unstructured text data (word2vec). The output is a vocabulary of words (which appear in the text) along with their representations in an n-dimensional vector space, and related words and/or groups of words appear close to each other in that vector space (Campr & Ježek, 2015). Training of this model is extremely efficient since it does not involve dense matrix multiplications. The created vectors encode many linguistic regularities and patterns, of which many can be represented as linear translations. The model can also be extended to creating phrase vectors instead of word vectors, to understand existing word-combinations that are different from the natural combination of the word's meanings. This makes the model more expressive and is relatively simple. After identifying phrases, they are treated as individual tokens during the training. Also, vector addition through basic mathematical operations (combinations of individual vectors) can often produce meaningful results in that a combination gets closer to related vectors. In following research, Le & Mikolov (2014) proposed a paragraph vector and document vector (doc2vec), which is an unsupervised algorithm that learns fixed-length feature representations from variable-length pieces of texts (sentences, paragraphs, and documents). It can be used to predict words in the document. The construction overcomes the weaknesses of bag-of-words models since it does not lose the word ordering and does not ignore semantics of words. Empirical results prove that it outperforms bag-of-words models and other techniques for text representations.

Menger, Scheepers, & Spruit (2018) compared bag-of-words with word embeddings (word vectors and paragraph vectors) in the context of predicting violent incidents based on Dutch clinical texts from a psychiatry department. The results showed that the deep learning techniques provide a relatively small but consistent performance improvement.

## 2.1.5  Dutch natural language processing

The English language is the most common language used in NLP tasks. The requirement in this research of processing Dutch questions and processing Dutch documents may pose additional challenges, limiting the NLP tasks of the to be designed QAS. In this paragraph, available NLP solutions for the Dutch language are discussed.

### *spaCy*
spaCy offers statistical models to predict and assign linguistic features. It has a processing pipeline consisting of a tagger, dependency parser, entity recognizer, text categorizer, tokenizer, lemmatizer, morphology, matcher, and phrase matcher (spaCy, 2018). For Dutch, a model trained the Universal Dependencies and WikiNER corpus, is available that assign context-specific token vectors, part-of-speech tags, dependency parse, and named entities. spaCy claims to achieve higher accuracy for named entity recognition compared to human annotations (spaCy, 2018).

### *Frog*
Research groups at the Tilburg University and the University of Antwerp have developed memory-based NLP modules for Dutch NLP since the 1990s. Over the years, these modules have been integrated into a single text processing tool called Frog, which is currently maintained and developed by the Radboud University Nijmegen (van der Sloot, van den Bosch, & van Gompel, 2018). Frog performs tokenization, part-of-speech tagging, lemmatization, and morphological segmentation of word tokens. At the sentence level, it identifies non-embedded phrase chunks in the sentence, recognizes named entities and assigns a dependency parse graph. In reflection to the QAS design and specifically questions and document analysis, Frog supports the desired morphological and syntactic analysis, and semantic analysis is also possible through part-of-speech tagging.

### *word2vec and doc2vec*
As stated before, question and answer embedding through vectors can be of use since it is relatively easy to implement while it can provide a semantic analysis of text very well, especially compared to more traditional methods. Embeddings can be done on multiple levels: word2vec for words (and phrases) (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013), and doc2vec for fixed-length texts (paragraphs and documents) (Le & Mikolov, 2014). This, to create vector neural networks to perform matching between embedded questions and answers. There are several word2vec models trained on Dutch corpora available, such as the Dutch embeddings used by Tulkens, Emmery, & Daelemans (2016) based on different Dutch corpora which are available on GitHub.

## 2.1.6  Quality of produced answers

Allam & Haggag (2012) present possible criteria for the evaluation of answers delivered by a QAS: *relevance*, *correctness*, *conciseness*, *completeness*, and *justification*. Based upon these criteria, answers can be judged to be correct (relevant and correct), inexact (not concise and not complete), or unsupported (no justification).

- *Relevance*: the answer should be a response to the question.
- *Correctness*: the answer should be factually correct.
- *Conciseness*: the answer should not contain extraneous or irrelevant information.
- *Completeness*: the answer should be complete (not a part of the answer).
- *Justification*: the answer should be supplied with enough context to allow a user to determine why this was chosen as an answer to the question.

Most commonly used measures that are typically utilized for automated evaluation of classification performance are *precision*, *recall*, and *F-measure*:

- *Precision*: the amount of answers given being correct (the number of correct answers divided by the number of questions answered).
- *Recall*: the amount of questions answered with correct answers (the number of correct answers divided by the number of questions to be answered).
- *F-measure*: a weighted average of precision and recall (2 * (precision * recall) divided by (precision + recall)).

## 2.1.7  Research towards the experience of a question-answering system

Research towards QASs is mainly focused on the underlying technology of the system, attempting to achieve more effective, efficient, and foremost accurate answering for a growing amount of questions. Improving existing technology, techniques, or approaches, or introducing new ones seems to have the highest priority in research. Answering questions is of course the main goal and without correct answers, a QAS would be worthless. However, very few studies are found which focus on the experience of using a QAS (let alone UX). Two studies are found that describe more user-centered QAS research.

### *Interface conditions: amount of context with answer*

Lin, et al. (2003) have investigated interfaces of QASs because they are largely unexplored. More precisely, their user study focused on the amount of context returned with the answer to a question. They proposed four types of so-called interface conditions: the 'exact answer' (only the exact answer), the 'answer-in-sentence' (the exact answer along with the sentence from which the answer was extracted), the 'answer-in-paragraph' (the paragraph from which the answer was extracted with the sentence containing the answer being highlighted), and the 'answer-in-document' (the entire document from which the answer was extracted with the sentence containing the answer being highlighted). The results of the study showed that the 'answer-in-paragraph' is preferred the most, and the 'exact answer' is the least preferred interface condition. Paragraphs were perceived as a good size chunk of information, while the exact answer was too little, and the entire document was often too much. The 'answer-in-sentence' often did not provide any useful additional information compared to the exact answer only. When a multiple-question scenario occurs, the completion time is not affected by the type of interface condition. However, the effect on the amount of questions required to complete the scenario was found very significant. Logically, the 'answer-in-document' required way less questions compared to the 'exact answer' interface condition. These findings are interesting since QASs (in comparison to IR) are developed to deliver answers rather than documents. Lin, et al. (2003) address that they revealed the schism between the technological drive and actual user preferences, and they believe that user considerations should be treated on an equal footing with the underlying technology.

### *User satisfaction with question-answering systems*

Ong, Day, & Hsu (2009) have developed an evaluation model for user satisfaction with QASs based on existing information system models and theories, incorporating constructs from both user satisfaction and technology acceptance literature. They address the little work on evaluation models for QASs despite the many implementations, while an appropriate evaluation leads to suggestions for overall QAS architecture and behavior improvements and could help to determine the extent to which the system meets its requirements. Ong, Day, & Hsu (2009) also address that most evaluation models are system-centered, meaning that that there is less

attention for user-centered evaluation while a practical QAS should satisfy its users. An evaluation model should provide feedback on the system's architecture but also on the impact of its behavior on the user, which thereby facilitates system improvements including improvements that benefit the user.

The model (or measurement instrument) consists of 18 items among four major constructs: ease of use (the degree to which a person believes that using a particular system would be free of effort), usefulness (the degree to which a person believes that using a particular system would enhance his or her job performance), service quality (the user's judgement of the overall excellence of a QAS based on assurance, empathy, and responsiveness), and information quality (the quality of the content). These constructs combined measure the user satisfaction of a user of a QAS, and a high satisfaction score would indicate that the user finds the QAS to be very easy to use, to be very useful, to have high quality of service, and to have high quality of information. The technical quality of the system is considered irrelevant since a technically superior system would not be considered successful if it does not the meet user's need. Therefore, this construct is absent in the model.

## 2.1.8 Summary

> ### *Question-answering system architecture*
> A fairly standard pipeline of three modules, each purposed for an accompanying phase of the whole question-answering process:
> - The question processing module is concerned with analyzing, classifying, and reformulating the question of the user.
> - The document processing module is concerned with retrieving, filtering, and ordering documents in which answers must be sought.
> - The answer processing module is concerned with identifying, extracting, and validating the answer to the question.

> ### *Examples of question-answering systems in practice*
> Examples of QASs applied in practice have been briefly described. This covered QASs in a:
> - Helpdesk context, based on patterns written by CSRs rather than application data, for a customer service purpose.
> - Clinical context, delivering question-focused extractive summaries, which has the potential to assist physicians and thereby improve the quality of patient care.
> - Physics domain context, to enable users to use natural language to retrieve information, which proved to be usable in real environments. It was built in a way to be applied in other domains and languages as well.

> ### *Classifications of question-answering systems*
> Existing variations of QASs have been discussed through multiple classifications.
> - The domain for which the QAS is purposed for and for which it is supposed to answer questions, which can be either an open domain or a restricted domain.
> - Types of questions a QAS strives to answer, which can be factoid, list, hypothetical, causal, confirmation, and opinion type.
> - Techniques to analyze the question from a user and analyze the content of a document, which can be morphological, syntactic, semantic, pragmatic and discourse, expected answer, and focus recognition analysis.
> - The structure of the data source, which can be structured, semi-structured, and unstructured.
> - Characteristics of the data source's content, which are the source size, language, heterogeneity, genre, and media.
> - A matching function is required to match the content of a question with a fitting piece of information from the document collections, which can be realized through a set theoretic, algebraic, probability, feature-based, or conceptual graph-based retrieval model.
> - The form in which answers are presented to the user, which can be an extracted answer or a generated answer.
> - The general technique used to achieve the overall QAS process of answering a question, which can be data mining, IR, NLU, or knowledge retrieval and discovery.

➢ *Other question-answering techniques*

Additional techniques have been explored that were not covered in the classification while possibly being of value. This resulted in finding FAQ answering through question similarity and a deep learning-based model for question-answering. The former uses existing FAQ (i.e. question-answer pairs) and matches user questions to the questions available in this set. The corresponding answer of a matched question is returned as answer. The latter uses question and answer embedding through vectors. It calculates similarity scores for question and answer pairs to indicate which answer suits the questions the most.

➢ *Dutch natural language processing*

To be able to process Dutch language in a QAS, available NLP solutions for Dutch language have been explored.

- spaCy offers a model trained on the Universal Dependencies and WikiNER corpus, is available that assign context-specific token vectors, part-of-speech tags, dependency parse, and named entities.
- Frog is a text processing tool that integrates memory-based NLP modules for Dutch NLP. Frog performs tokenization, part-of-speech tagging, lemmatization, and morphological segmentation of word tokens. At the sentence level, it identifies non-embedded phrase chunks in the sentence, recognizes named entities and assigns a dependency parse graph.
- word2vec (words and phrases) and doc2vec (paragraphs and documents) can be used to create vector neural networks to perform matching between embedded questions and answers. Models trained on Dutch corpora exist.

➢ *Quality of produced answers*

The quality of answers delivered by a QAS can be evaluated based on several criteria: relevance, correctness, conciseness, completeness, and justification. For classification performance, the precision, recall, and F-measure can be used.

➢ *Research towards the experience of a question-answering system*

Research towards QASs is mainly focused on the technology, attempting to achieve more effective, efficient, and accurate answering for a growing amount of questions. Knowledge on the experience of using a QAS is lacking. Two studies are found that describe more user-centered QAS research.

- An interface condition has been researched, specifically, the amount of context provided with the answer. It showed that providing a whole paragraph with the answer being highlighted is preferred the most.
- An evaluation model for user satisfaction with QASs is developed. An appropriate evaluation leads to suggestions for improvements. While most evaluations are system-centered, this model is user-centered by measuring ease of use, usefulness, service quality, and information quality. The model intentionally does not cover technical quality.

## 2.2    User experience

This section discusses UX. First, UX is defined as the field of study to be studied for this research by reviewing multiple fields of study related to this research. After that, it is discussed what separates UX from former knowledge on people's interaction with technology. Next, UX is defined and it is explained what is interpreted by its definition. This is followed by an explanation of the key elements that are considered to encapsulate how UX unfolds. The next paragraph discusses research which resulted in insights about the elements that play an essential role in sparking positive UXs. Finally, principles to design UXs are discussed. The section ends with a summary of the discussed content.

### 2.2.1    Concepts related to user experience

While this research focuses on UX in relation to a QAS, it is helpful to address several fields of study considered relevant in a research towards the experience of a customer. This, to position those fields of study and to argue for UX as the central field of study for this research. The concepts related to the experience of customers are customer experience, customer service, self-service technology (SST), and human-computer interaction (HCI). A sub-literature review covering these concepts can be found in Appendix A: Concepts related to the experience of customers. This sub-literature review first discusses customer experience, which focuses on the experience of all interactions between a customer and a company. This encompasses the experience of a customer in its broadest sense, also described as the total experience. Next, customer service is defined, and it is described what is perceived as an approach that satisfies and dissatisfies customers. This is followed by a discussion of SST which is technology that enables customers to fulfill customer services themselves through an interface. Again, satisfiers and dissatisfiers are described. Finally, HCI is discussed which involves designing, implementing, and evaluating interactive systems in the context of user's task and work that are effective, efficient, easy, and enjoyable to use. The sub-literature review ends with a summarization of insights that can be derived from the discussed concepts.

Figure 3 presents an overview of these concepts (dashed, white boxes) and the entities; a customer who can interact with an organization or a system of the organization (solid, grey boxes). It loosely visualizes the positions and mutual relationships between the concepts and entities. The smaller the box, the more specific the concept is. Customer experience is broader than just the organization alone (i.e. not fully controlled by the organization). The customer can interact with the organization, partly causing a certain customer experience. Customer service represents a part of the organization with which the customer can interact. For customer service, (a) certain system(s) can be made available to customers to interact with. This enables the existence of SST as a provider of customer service, for example, a QAS through which a customer can utilize a service independently (i.e. getting answers to questions). Such SSTs can be subject to HCI theory. Within the field of HCI, UX is the concept specifically concerned with the experience someone has when interacting with a specific system (e.g. a QAS). Because of this, it provides the most direct insights on enhancing the experience for customers that are offered to use a QAS, more than any of the other concepts
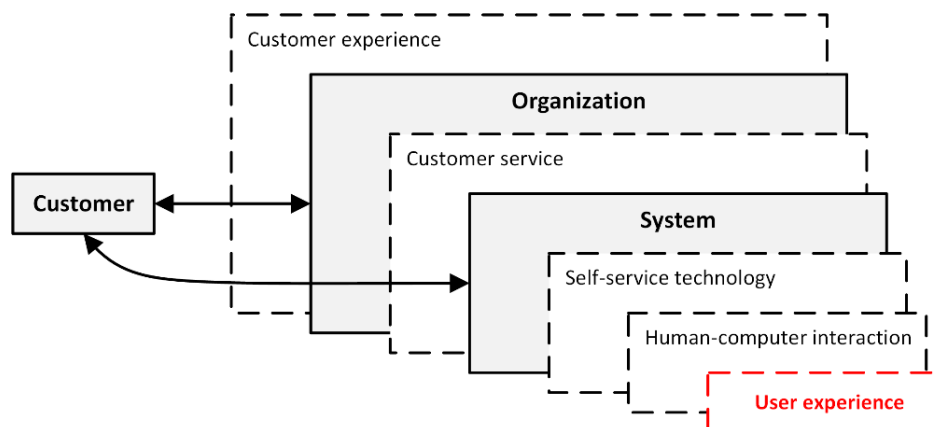


*Figure 3. Positioning of concepts related to user experience.*

## 2.2.2 From functionality to experiences

Hassenzahl (2008) states that UX itself is a truly extended and distinct perspective on the quality of interactive technology. It shifts the focus from products and problems to humans and the drivers of positive experiences. Only looking at the accomplishment of tasks with interactive technology is too short-sighted when analyzing what people do with and gain from technology usage. In industry, it is common to treat UX as the product's usability and cover it with user-centered design. However, it is the underlying motives for technology use, such as insight, pleasurable stimulation, and social exchange, that generate true outcomes of technology use: feelings and experiences.

UX finds its roots in the field of HCI research. The HCI community seemed to acknowledge that functionality and usability alone are not enough to understand the relationship between a user and a product (Hassenzahl, 2003). UX provides an additional way of looking at this relationship, but what exactly is UX? Hassenzahl & Tractinsky (2006) derived three distinctive, prominent perspectives on UX, based on its presence in HCI research: (1) beyond the instrumental, (2) affect and emotion, and (3) the experiential. Each perspective contributes a facet to the understanding of people's interaction with technology.

1. Ensuring the product's instrumental value has been the major endeavor of the HCI field, because the user's task has been the key point of user-centered analysis and evaluation techniques. However, a product's beauty (or aesthetics) was identified as an important quality aspect of technology because of its intrinsic value: beauty satisfies a general human need. In addition, for a whole range of non-instrumental needs (e.g. surprise, diversion, intimacy), it is argued to be of importance to be addressed in technology. Therefore, system designs should not only be concerned with a product's pragmatic aspects (i.e. it's fit to the user's behavioral goals) but also with hedonic aspects (i.e. it's fit to the user's self). Thus, the product's quality should also be based on non-instrumental aspects for a more complete holistic system design.

2. The importance of an affective system is emphasized by HCI research for a wide range of central processes (e.g. human decision-making, subjective wellbeing), emphasizing that it is important to consider the user's emotions and affect during technology interaction. The Affective Computing project attempted to address affect but took the computer's perspective by predominantly dealing with how to sense user affects in order to adapt to it or express an affective response. Since the human interaction with technology mostly results in negative emotions, Affective Computing was focused on mechanisms to detect and undo those. UX shares the recognition of emotions and affect, but instead, it is concerned with affective consequences from the human's perspective. The goal is to understand the role of affect as an antecedent, a consequence, and a mediator of technology use. The focus is on enabling positive emotions (e.g. joy, fun, pride) and use that as the core objective. This is a shift from the usual core objective in HCI (even from its most cognitively driven perspective) of preventing frustration and dissatisfaction. UX is also found in Technology Acceptance literature, which studies the interplay between user perceptions (e.g. perceived usability) and the intention to use or the actual behavior. Research (on the use of a course management system) provided evidence that perceived affective quality is an antecedent of the system's perceived usability, its usefulness, and its intention to use.

3. The experiential perspective on UX emphasizes the situatedness of technology use and the temporality of technology use. This perspective views UX is a unique combination of elements such as the product and the internal states of the user (e.g. mood, expectations, active goals) which extends over time with a definitive beginning and end. The elements are assumed to be interrelated, meaning that they interact and modify each other. The outcome of this process is the actual experience. In contrast to material outcomes, the advantage of experiences is that they have a more positive impact on a person's wellbeing: they possess affective quality and help to transform and regulate affective states.

In Figure 4, the perspectives are visualized as components of what is understood to distinguish UX within the HCI research field. Short statements are formulated based on the understanding of each perspective, which define how UX goes beyond functionality and usability.
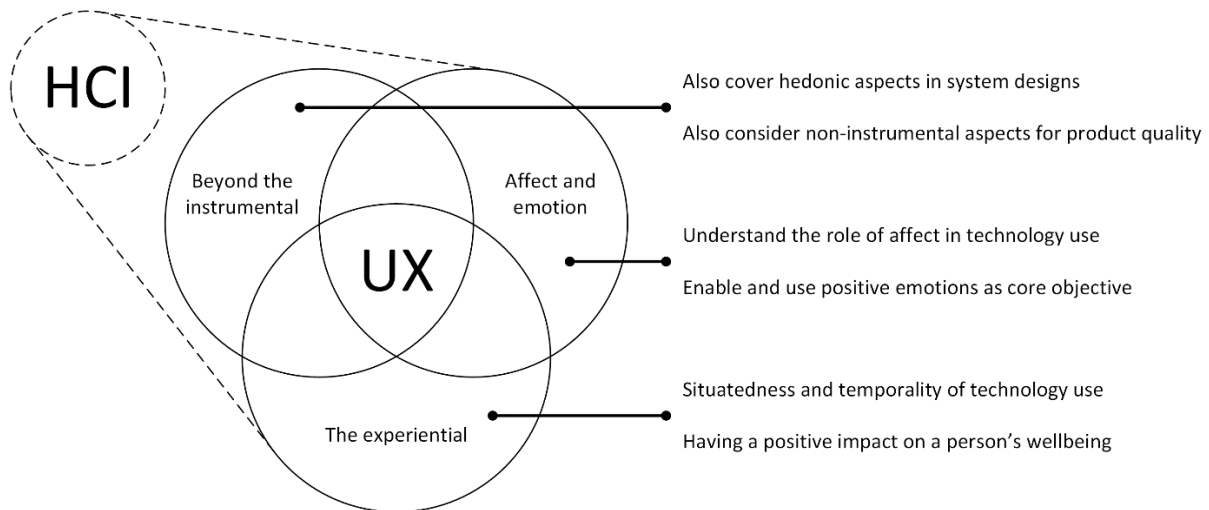
*Figure 4. User experience from an HCI perspective. Adapted from Hassenzahl & Tractinsky (2006).*

Based on these perspectives, it can be stated that UX is about technology which fulfills more than just instrumental needs by acknowledging its use as a subjective, complex, and dynamic encounter. The UX is seen as a consequence of the characteristics of the designed system, a user's internal state, and the usage situation in which the interaction takes place.

## 2.2.3   Defining user experience

Hassenzahl & Tractinsky (2006) only indicated which perspectives within HCI research are of relevance for understanding people's interaction with technology that are beyond functionality and usability (and therefore proposed as facets of UX). To understand what is understood by the term UX, definitions of UX from literature are discussed. Before defining 'user experience', we first explore the concept of 'experience' itself.

### *Experience*

According to Hassenzahl (2008), experience itself is an ongoing reflection on events that people currently go through or a constant stream of self-talk. One essential element of experience is the momentary feeling of pleasure and pain in various intensities (good-bad), which regulates behavior. People can question themselves whether they feel good or bad during an event: a 'psychological currency' which allows to qualitatively compare experiences. This is the central basis of subjective product evaluation. Hassenzahl, Diefenbach, & Göritz (2010) agree by stating that experience is a stream of feelings, thoughts, and actions; a continuous commentary on the self's current state of affairs. Experiences are characterized as ubiquitous, mostly unconscious, but accessible to the person experiencing, and people can summarize and memorize particularly outstanding, rich, or touching experiences. Roto, Law, Vermeeren, & Hoonhout (2011) provide a similar but clearer, compact definition: experience is the individual's stream of perceptions, interpretations of those perceptions, and resulting emotions. They add that this experience may be different for each person, which emphasizes the individual and dynamic nature of experiencing. Figure 5 shows the interpretation of experience through a visualization.
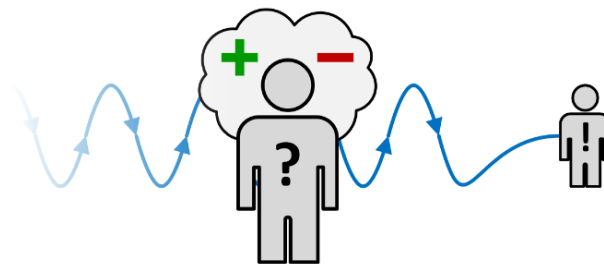


*Figure 5. Interpretation of experience visualized.*

The blue line represents the event which is experienced by an individual. During this event, the individual internally processes the experience of the event, e.g. an ongoing reflection, self-talk; a stream of feelings, thoughts, and actions; commentary on the self's current state of affairs; a stream of perceptions, interpretations

of those perceptions, and resulting emotions (represented by the thought cloud with outcomes that can be positive and negative). This is a subjective product evaluation that is specific to the individual; the experience can therefore be perceived differently per individual, which emphasizes the dynamicity of experience (represented by the question mark). Finally, an overall perception of the experience is formed by the individual (represented by the exclamation mark). The experience is now accessible for the individual and can be summarized and memorized.

### User experience

Various definitions for UX can be found in literature, but they are fairly in line with each other. The ISO definition and the three most prominent views on UX are discussed. The ISO definition is "a person's perceptions and responses that result from the use and/or anticipated use of a product, system, or service" (ISO 9241-210:2010, 2010). This definition emphasizes the subjectivity of UX, by focusing on a person's immediate consequences (perceptions and responses) of use and specifically mentioning anticipated use as well. Hassenzahl (2008) explicitly addresses the subjectivity side of products as well. He defines UX as a momentary, primarily evaluative feeling (good-bad) while interacting with a product or service. In the context of interactive products, these interactions are by technology mediated human-product interactions, which stretch over time and have a definite beginning and end. He addresses that with UX, the focus shifts from products and materials (e.g. content, function, presentation, interaction) to humans and feelings: the subjective side of product use. Additionally, dynamicity of the user is emphasized (e.g. motivation for usage, mood, mental and physical resources, expectations of usage). UX is seen as a temporal phenomenon, that is present-oriented and changing over time. While retrospective or prospective judgements about experiences are not neglected, judgements are primarily based on the stream of passing momentary feelings. Law, Roto, Hassenzahl, Vermeeren, & Kort (2009) recommend scoping the term UX to products, systems, services, and objects that an individual interacts with through a user interface. Face-to-face interaction between humans is not considered to be in this scope. Based on their conducted survey to collect views from academia and industry on UX to understand, scope, and define UX, they summarize UX also as being subjective and dynamic but context-dependent as well. Context factors are assumed to influence the UX. More specifically, it is stated that the UX is affected by the current context, in addition to the internal state of the individual and earlier experience. This also emphasizes that the UX is present-oriented. Furthermore, it is stated that feelings and experiences unfold per individual (even when a group experiences something together, it is still an experience per individual). Roto, Law, Vermeeren, & Hoonhout (2011) define UX as an encounter with a system that has a beginning and end as well and put focus on the overall designation of how people have experienced the system for a certain period. This view emphasizes the outcome and memories of an experience rather than its dynamic nature. They also address context as a main category of influencing factors: social context (e.g. involvement of others, being surrounded by others), physical context (e.g. sitting on the couch, walking outside), and task context (i.e. tasks surrounding the main task). Figure 6 shows the interpretation of UX through a visualization.
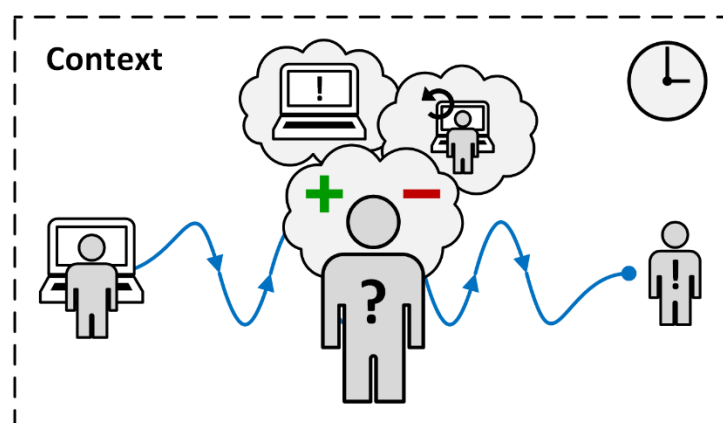


*Figure 6. Interpretation of user experience visualized.*

The visualization is similar to that of experience but more specific. The blue line represents the experience of an event, which is experienced by an individual. For UX, the event is specifically a by technology mediated interaction; an interaction with a product, system, or service through a user interface (represented by the laptop and user combination). The experience stretches over time with a definite beginning (represented by the blue line starting at the laptop and user combination) and end (represented by the dot at the end of the blue line),

after which the experience becomes a memory of the individual (represented by the exclamation mark). This emphasizes that the experience is primarily present-oriented (represented by the clock). Judgements primarily based on the individual's stream of momentary feelings during the performed interaction form the experience; the individual's internal processing (represented by the main thought cloud with outcomes that can be positive and negative). Anticipated use also plays a role. This is structured by expectations of the product (represented by the thought cloud with the laptop with exclamation mark) and previous experiences with the product or similar products (represented by the thought cloud with the laptop and user combination with the counterclockwise arrow). The evaluative feeling resulting from this is specific to the user; the UX is therefore perceived differently per user (represented by the question mark). Finally, contextual factors in relation to the interaction also influence the experience (represented by the dashed frame). All these elements, assumed to form the UX, portray how UX is about the subjective side of product use.

## 2.2.4  Key elements of user experience

Hassenzahl (2003) attempted to define key elements of UX and their functional relations, aimed at addressing the subjective nature of experience, perception of a product, and emotional responses to products, in varying contexts. Hassenzahl (2003) proposed a model more complex compared to models for product functionality and usability, to highlight key elements of UX. A simplified and slightly adjusted version of the model, based on further research by Hassenzahl (2008), is shown in Figure 7. The figure shows a designer perspective and a user perspective. In short, the model expresses that from a designer perspective, a product has product features that result in a product character as intended by the designer. Such a product character consists of pragmatic attributes and hedonic attributes. From a user perspective, this results in a perceived product character based on its perceived pragmatic quality and hedonic quality. Finally, consequences of using the product are dependent of the perceived product character and the context in which the product is used. These consequences define the user's feelings about the product.
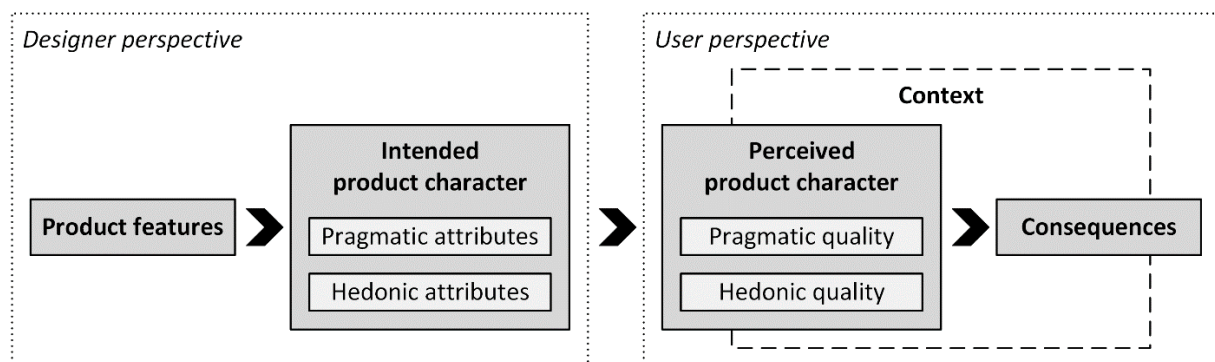


*Figure 7. Model of the key elements of user experience. Adapted from Hassenzahl (2003).*

### *Product character*
The product character is a high-level description and summarizes a product's attributes such as interesting, useful, or predictable (Hassenzahl, 2003). Its function is to reduce cognitive complexity and to trigger the user into handling the product in a certain way. The designer chooses and combines product features: its content, presentational style, functionality, and interactional style. Thereby, the designer conveys a product character. However, this character is subjective and there is no guarantee that the user will perceive and appreciate the product in the way the designer intended it. A user perceives a personal version of the product character. This is based on the combination of the product's features and the individual's personal standards and expectations (e.g. comparison to other products). The latter explains the variations between individuals and the possibility of a change in apparent product character for an individual (e.g. due to familiarity with the product). Subsequently, this leads to consequences based on the fit between the perceived product character and the individual's context in which the product is used: the individual's emotional consequences and a judgement about the product's momentary appealingness. This also means that attributes of a product character become more or less relevant for the consequences depending on the context, and consequences can vary strongly because they are embedded in a particular context. This process of perceiving and constructing the character and experiencing consequences will always take place. The outcome may change with growing knowledge of and experience with the product, implying that this process repeats constantly. For example, an ATM shows its user step-by-step what

happens with every action to make the money withdrawal process highly understandable. When someone uses this ATM for the first time, it can be praised for this and be perceived satisfying. However, when someone is already familiar with using this ATM, its highly informative character can be perceived rather annoying and inefficient, leading to frustration.

### *Pragmatic and hedonic attributes and pragmatic and hedonic quality*
The intended product character is a structure of pragmatic attributes and hedonic attributes.
- ➢ *Pragmatic attributes* are meant to enable the user to manipulate the environment. This requires pragmatic attributes for relevant functionality and to access this functionality (i.e. usability). For software products, typical pragmatic attributes are 'supporting', 'useful', and 'controllable'. A product that is perceived as pragmatic is primarily instrumental and is used to fulfill behavioral goals, or so-called do-goals (e.g. booking a hotel, finding a song in a music-library), that are externally given or internally generated.
- ➢ *Hedonic attributes* are all product attributes that are not classified as pragmatic attributes. Instead of behavioral goals, hedonic attributes emphasize the individual's psychological well-being. 'Hedonic' expresses the belief that attributes it subsumes are strong potentials for pleasure, especially compared to pragmatic attributes. Typical hedonic attributes are 'impressive', 'exciting', and 'interesting'. Therefore, a product that is perceived as hedonic is used to fulfill basic human needs beyond the instrumental, or so-called be-goals (e.g. need for novelty, personal growth, self-expression, relatedness).

Hassenzahl (2008) argues that people's perceptions of products are defined by two product quality dimensions, based on the described product attributes: pragmatic quality and hedonic quality. Therefore, it is chosen to define the perceived product character as a structure of pragmatic quality and hedonic quality.
- ➢ *Pragmatic quality* is defined as the product's perceived ability to support the achievement of do-goals, which is therefore defined by the product's perceived functionality and usability in relation to tasks of the user (i.e. do-goals).
- ➢ *Hedonic quality* is defined as the product's perceived ability to support the achievement of be-goals, which is therefore defined by the user's self; the perceived fulfillment of the user's human needs (i.e. be-goals).

### *Consequences*
Experiencing a product with a certain product character has emotional consequences: satisfaction and pleasure (Hassenzahl, 2003). These consequences are momentary and related to the usage situation (i.e. context). Satisfaction depends on the match between the expectations of using a particular product and the usage outcome. If expectations are confirmed, the user will feel satisfied. In contrast, pleasure does not require expectations. Pleasure is being pleased about a desirable event. The more unexpected the event is, the more intense the pleasure will be, and when a particular product's experience has a desirable deviation from the product's expectations, it will be pleasing. Appealingness of a product depends on the ability of a product to trigger positive emotional reactions. Contributing product attributes are, for example, 'sympathetic', 'attractive', and 'motivating'. Perceptions of such product attributes are weighted and integrated for a product's appealingness by considering the user's particular context. It means that experiences with and feelings towards a product in a particular situation are integrated into an evaluative judgement to define its appealingness.

## 2.2.5 Enabling positive user experience

The previous section discussed how UX unfolds. The question is how a positive UX is enabled. Research that specifically looked at needs and affection through product use is discussed in this section, to uncover what is assumed to lead to positive UX.

### *Human needs in relation to affective states*
Hassenzahl (2008) defines a good UX as the consequence of fulfilling human needs for autonomy, competency, stimulation (self-oriented), relatedness, and popularity (other-oriented) through interacting with the product or service. Thus, positive UX is dependent of the product's hedonic quality because that quality directly contributes to the core of positive experiences. The product's pragmatic quality is facilitating users in fulfilling be-goals and is therefore only indirectly contributing to positive experiences by making that process more easy and likely. Thus, usability itself is not of value, but it gets value by facilitating the process of fulfilling meaningful be-goals. The positive experience is 'made' through fulfilling a basic human need, i.e. psychological need. This fulfillment

is the cause for positive emotions and pleasurable experiences in product interaction and leads to the association of hedonic attributes to the product. It can be discussed whether people have a primary intention of 'feeling good' or 'fulfilling the need of being competent' when they start an interaction with technology. Therefore, fulfillment of be-goals may be rather implicit and automatic for a person. The positive feeling that a person experiences is more graspable than the actual reasons for it, and this leads to believe that people have expectations for experiences instead of explicit be-goals.

To find proof for these claims, Hassenzahl (2008) conducted an exploratory study to clarify what makes an experience with technology satisfying. Participants were asked to fill in a questionnaire about a satisfying experience with technology. First, the three core human needs, according to the Self-Determination Theory, were measured: autonomy, competence, and relatedness. Second, feelings that participants had during the satisfying experience were measured by using the Positive And Negative Affect Schedule (PANAS) (Watson, Clark, & Tellegen, 1988). This, in order to correlate the saliency of human needs with the intensity of affective states (i.e. feeling inspired, alert, excited, enthusiastic, determined, afraid, upset, nervous, scared, and distressed). The first results showed that competence is the most salient need, followed by autonomy, and lastly relatedness. Secondly, calculated correlations showed that autonomy is solely linked to positive affect and specifically with feeling inspired. Competence showed a similar pattern, but its experience was also accompanied with feeling nervous and scared. Relatedness was not significantly correlated to positive affect. It actually uncovered that a feeling of relatedness was accompanied by a feeling of distress. In conclusion, this exploratory study showed that autonomy and competence are a source for positive experience through technology usage. Relatedness played a minor role and is not related to positive affect, which could be caused by the lack of addressing social needs in the technologies' designs at the time. Another interesting find was that different needs produce different qualities of affective experience. The need competence leads to activation (positive) but also to fear of failure (negative). Lastly, an important observation was that none of the core human needs (according to the Self-Determination Theory) leads to excitement and enthusiasm, while those are interpreted as the qualities most related to positive UX, specifically the emotional consequence pleasure.

### *Human need fulfillment as a cause for positive experience*
Sheldon, Elliot, Kim, & Kasser (2001) explored the relationship between human needs and affect. A list of top ten human (or psychological) needs was compiled to asses need fulfillment, and the PANAS was used to assess affect. Three main findings resulted from this: (1) the degree of need fulfillment was positively related to the intensity of positive affect, (2) needs were relatively independent from each other which means that satisfying experiences can be classified (because they are marked by particular needs), and (3) needs differed in their saliencies, with autonomy, competence, and relatedness most notable in reported positive life events. In conclusion, fulfillment of particular psychological needs is a source of positive experience. This causal relation is however an assumption, but it has been shown that positive experiences can be distinguished based on the primary need fulfilled. Hassenzahl, Diefenbach, & Göritz (2010) recognized that assumption in particular models of UX. They compared the models with the list of top ten human needs, and to a certain extent, the needs competence, relatedness, popularity, stimulation, and autonomy are found in all models (in various wordings). From that, it can be concluded that these needs further clarify the psychological basis for assumed pleasures, needs, or values. Hassenzahl, Diefenbach, & Göritz (2010) conducted a study to broaden the understanding of pleasurable experiences with technology, from this standpoint that the major cause for positive experiences with interactive technologies is fulfillment of psychological needs. In their approach, they first questioned what these needs exactly are. They made a selection of seven out of the ten human needs which they consider most important in the context of experiences with technology. The first five are derived from the comparison (i.e. competence, relatedness, popularity, stimulation, and autonomy) and those are complemented with security (due to its link to usability issues) and autonomy (due to its central character in the Self-Determination Theory). However, a principal component analysis led to the conclusion that 'meaning' and 'autonomy' are largely loaded on the same component. Since 'autonomy' appears more frequently in UX literature compared to 'meaning', it is chosen to use 'autonomy' and its description to cover both needs. The six human needs and their descriptions can be found in Table 5.

*Table 5. List of the six human needs and their description considered to be most important in the context of experiences with technology (in alphabetic order). Adapted from Sheldon, Elliot, Kim, & Kasser (2001) and Hassenzahl, Diefenbach, & Göritz (2010).*

| Human need | Description |
|---|---|
| Autonomy | Feeling like you are the cause of your own actions rather than feeling that external forces or pressure are the cause of your action. |
| Competence | Feeling that you are very capable and effective in your actions rather than feeling incompetent or ineffective. |
| Popularity | Feeling that you are liked, respected, and have influence over others rather than feeling like a person whose advice or opinion nobody is interested in. |
| Relatedness | Feeling that you have regular intimate contact with people who care about you rather than feeling lonely and uncared of. |
| Security | Feeling safe and in control of your life rather than feeling uncertain and threatened by your circumstances. |
| Stimulation | Feeling that you get plenty of enjoyment and pleasure rather than feeling bored and under stimulated by life. |

It is emphasized that this is not a definitive list of needs, but it can be used to cover most experiences without overcomplicating things. To explore the role of need fulfillment for a user's experience, a questionnaire study has been conducted. Participants were asked to recall and relive a recent, outstanding positive experience with technology in order to rate it on four scales: (1) need fulfillment, (2) experienced affect, (3) attribution (i.e. to what extent the product causes the experience), and (4) product perception and evaluation. The results of this study provide insights in (1) types of experiences, (2) saliency of needs, (3) affect in relation to need fulfillment and product perception and evaluation, and finally (4) need fulfillment, positive affect, and attribution in relation to pragmatic quality and hedonic quality.

1. Various experiences through various interactive products and activities were identified in the study, but for a classification of the content of experiences, only the distinction between social (37%) and non-social (63%) was made.

2. The most salient needs of positive experiences are relatedness, stimulation, and competence, followed by less salient needs (in descending order) popularity, security, and autonomy. From these results, it can be stated that technology is experienced positive if it facilitates closeness and communication between people, provides new stimulating insights, and provides opportunities for mastery. Relatedness and popularity were more salient in social contexts but only relatedness was significantly more salient. The other needs were all negatively correlated (significantly, except for stimulation) which suggest that fulfillment of a particular need in a particular context (e.g. relatedness in a social context) hinders the fulfillment of others.

3. All needs were significantly correlated to positive affect, with a rank order (i.e. stimulation, competence, relatedness, popularity, autonomy, security) similar to the rank order of salience of needs. The fact that need fulfillment was related to positive affect can be translated to need fulfillment being a source of pleasure. Stimulation, relatedness, competence, and popularity were found as the most influential needs. The results also demonstrated differences between experiences, which is two-fold: different types of experience can revolve around a particular need, and different types of experience can provide a particular set of emotions attached to the context and actions (e.g. the bitter-sweet experience of competence, strongly tied to goal-oriented behavior and challenge). Product perceptions and evaluation are obviously related to the experienced affect. Hedonic quality was more strongly related to positive affect compared to pragmatic quality, which supports the idea of a product's hedonic quality being responsible for creating a positive experience (the so-called motivator). The product's pragmatic quality can be seen as the enabler for fulfillment of needs by making things easy to diminish negative affect instead of being a direct source of positive experience (the so-called hygiene factor).

4. Multiple models were created to analyze the roles of need fulfillment, positive affect, and attribution in relation to pragmatic quality and hedonic quality. The results showed that there is a direct relation between needs and product perception in the case of hedonic quality only, and that this relation is moderated by attribution (i.e. the extent to which the product was perceived to be responsible for need fulfillment). This leads to the conclusion that a product's hedonic quality is determined by need fulfillment attributed to the product. For pragmatic quality, only a small correlation with need fulfillment was found, which was fully moderated by positive affect. The correlation is therefore assumed to be caused by a so-called halo effect which means that the positive affect influenced the perception of the

pragmatic quality. The results also showed the existence of indirect effects of need fulfillment on both pragmatic quality and hedonic quality via positive affect. For hedonic quality, this indirect effect was moderated by attribution, and the more the product was perceived to be responsible for the need fulfillment, the stronger the indirect effect became. From these findings, it is concluded that positive affect is a main outcome of need fulfillment, making it a legitimate predictor of hedonic quality.

In a similar, more recent research (Hassenzahl, Wiklund-Engblom, Bengs, Hägglund, & Diefenbach, 2015), positive affect and fulfillment of psychological needs were also proofed to be related to each other. While positive affect can be assumed to be the consequence of fulfilled needs, only a correlation between the two exists instead of causality. Again, their relationships with product perceptions were moderated by attribution. If this attribution is crucial, it is of interest to find out how to ensure that users attribute them to the product. The importance of evaluating products experientially is therefore emphasized, since product-oriented measures may be misleading. While usability is a necessary precondition for positive experiences, the positivity itself is the result of hedonic quality and should therefore be measured. It provides insights beyond usability. Lastly, it is claimed that only focusing on the experience can be just as misleading as the current practice of focusing only on the product; focusing on the combination has the most advantages.

While certain human needs are considered more salient, Hassenzahl (2010) argues that certain needs are not necessarily more important than others. Each need should cover an important class of experiences that is central to being human. Research by Sheldon, Elliot, Kim, & Kasser (2001) proofed that fulfilling needs that are personally important do not make a person happier compared to fulfilling personally unimportant needs. Human needs are interpreted as universal, important, and meaningful to any human. However, people can have a specific urge for fulfillment of a certain human need. It was also proofed that the presence of a certain human need might not play an important role in a positive experience, although its absence could become a major aspect of a negative experience. Hassenzahl (2010) states that it is important to identify situations in which the urge for a certain need is an important starting point for designing experiences. However, the true challenge is to fulfill all human needs without making this too obvious.

### *Human need fulfillment in relation to question-answering*

Clarity about the relationship between (automated) question-answering and UX is lacking. The fulfillment of human needs has been researched in relation to, for example, an automotive context (Korber, Eichinger, Bengler, & Olaverri-Monreal, 2013), media consumption (i.e. watching, listening, and playing) (Hassenzahl, Wiklund-Engblom, Bengs, Hägglund, & Diefenbach, 2015), social media usage (Zhu & Chen, 2015), and gamification elements (Zimmerling, Höllig, Sandner, & Welpe, 2019). However, there is no evidence of true means to create for instance relatedness. It is not explained how human needs can be addressed in a question-answering process. It is often a process of coming up with ideas as a system designer or use existing products or systems, and then analyze the presence of human needs and/or fulfillment of them. For example, in the automotive context, three representations about an interaction with a car were shown to participants of a survey. The first was a story describing a holiday trip with the family with new in-vehicle technology for passengers to interact with. This story was meant to evoke a feeling of relatedness, stimulation, and competence. The second was a story about a man driving home from work using multiple devices to reach that goal. This was meant to evoke a feeling of security and autonomy. The third was a commercial advertising a car, meant to evoke popularity. This survey approach shows that it is an interpretation of the designer on how to convey a certain need in a product.

## 2.2.6   Designing user experience

Hassenzahl (2013) presented a simple conceptual model for the process of designing experience, consisting of three levels: the *what*, the *how*, and the *why* of a product (see Figure 8). These levels should be considered when designing technology-mediated experiences.
1. The *what* addresses the things people can do through an interactive product (i.e. functionality), which is often closely related to the technology itself or the product genre.
2. The *how* addresses how to act with the product on an operational, sensory-motor level, which is even more closely related to the actual product that is to be designed and the context in which it is to be used. This can be considered as interaction design, which is meant to make the functionality accessible in an aesthetically pleasing way (i.e. usability).
3. The *why* addresses why people engage in activities: the underlying need(s) for which fulfillment is desired (i.e. human need fulfillment).

Hassenzahl (2013) is of the opinion that the first two combined are typically considered as the product. Furthermore, by providing an interaction with the product that is sensual, aesthetic, novel, or arranged to provoke a stimulating interaction, it is assumed to make the product experiential. However, this disregards people's deeper motivations to use a product. For example, a text message is not necessarily primarily a text. It could be a love message which fulfills the need for relatedness. Therefore, the conceptual model for experience has the third level: the *why* of product use. This *why* is closest to the self (i.e. the user who has internal needs), and the *how* is closest to world (i.e. the usage context which sets demands for the product). The combination of the *what* and *how* represents the product (i.e. the functionality in relation to its usage), and the combination of the *what* and *why* represents the experience (i.e. the functionality in relation to the fulfillment of needs).
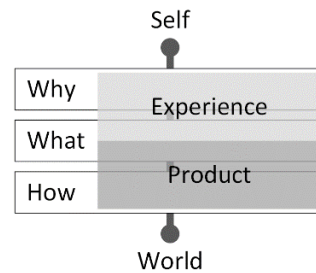


*Figure 8. Conceptual model for experience: three levels to consider when designing technology-mediated experiences. Adapted from Hassenzahl (2013).*

In other research, certain principles are found in other research that should be considered in the process of designing UX as well. They will be briefly discussed. Of course, an understanding of the underlying needs and a way to address them is required. Also, taking contexts and use cases in consideration during the process is beneficial. Lastly, system representations are required to evaluate ideas and explain their effects at an early stage.

### Understand and address underlying needs

Hassenzahl (2008) states that an interactive product should enable the user to accomplish do-goals, but that the product remains pallid if it does not have hedonic quality. When designing a product, its UX is more conceptual than designing something useful that is also good-looking. For example, a search functionality should allow discovery in a stimulating way instead of being a simple keyword-directed. With the design of a product, underlying be-goals should be considered and be repeated throughout the design process to create the experience. However, when the UX perspective is the true objective in a product's design, the underlying needs are to be taken seriously and require the design of technologies that match those needs. Particular techniques may be required for the fulfillment of particular be-goals. UX is focused on the psychological well-being of people and not on the performance as the outcome of human-product interaction.

### Consider contexts and use cases

In product design, potential consequences should be separated from actual consequences, because the former is more stable and therefore more reliable (Hassenzahl, 2003). This is argued for because emotional reactions and the judgement of appealingness (due to the perceived product character) vary due to different contexts. This means that a product design should focus on the product character and the context: it is important to understand why users judge a particular product as appealing, satisfying, and pleasing. Designing a product therefore requires a detailed understanding of the people and the context it is designed for (Hassenzahl, 2008). Since UX is strongly affected by contextual factors, Law, Roto, Hassenzahl, Vermeeren, & Kort (2009) suggest imagining use cases and their contextual factors, already during early development phases, in order to evoke realistic experiences.

### Develop system representations

Roto, Law, Vermeeren, & Hoonhout (2011) argue that representations of the system's concepts and designs need to be developed to give people a sense of what the experience might be like before the actual system is designed. With representations, feedback can be gathered to give direction for the design. Emotional responses should be captured, accompanied by explanations for them.

## 2.2.7 Summary

➢ *Concepts related to user experience*
Concepts related to UX are discussed in a sub-literature review because of their relevance to this research. Customer experience, customer service, SST, and HCI are discussed but were placed outside of the research scope. UX is the only concept specifically concerned with the experience someone has when interacting with a specific system and is therefore providing the most direct insights on enhancing the experience for customers that are offered to use a QAS.

➢ *From functionality to experiences*
The HCI community acknowledges that functionality and usability alone are not enough to understand the relationship between a user and a product. UX is about technology that fulfills more than just instrumental needs. Based on HCI research, UX is a distinctive field of study for the following reasons.
- Hedonic product aspects (non-instrumental product aspects) are also responsible for a product's quality, which is why system designs should also cover them besides pragmatic product aspects.
- Users' emotions and affect during technology interaction are important to consider. The goal is to understand the role of affect as an antecedent, a consequence, and a mediator of technology use. This, to enable positive emotions as a core objective, rather than preventing frustration and dissatisfaction.
- Technology use has a situatedness and temporality, resulting in a unique combination of elements that are assumed to be interrelated. The outcome is the experience, and the advantage of an experience is that is has a more positive impact on a person's wellbeing.

➢ *Defining user experience*
Definitions of experience and UX are discussed to fully understand and interpret the term.
- Experience is an ongoing reflection on events that people currently go through, a constant stream of self-talk, leading to certain emotions. It is a momentary feeling of pleasure and pain in various intensities, regulating behavior. This enables people to compare experiences, the central basis of product evaluation. Experiences are mostly unconscious but the person experiencing can summarize and memorize them. The experience is specific to the person experiencing, which emphasizes that it has an individual and dynamic nature.
- UX is a person's perceptions and responses that result from the use and/or anticipated use of a product, system, or service. It is a momentary, primarily evaluative feeling (good-bad) of a by technology mediated human-product interaction (through a user interface) with a product or service. The subjective side of product use is the core. UX is present-oriented and changing over time, but judgements are primarily based on the stream of passing momentary feelings. It is also context-dependent meaning that foremost the current context is assumed to influence the UX. It also has an individual and dynamic nature.

➢ *Key elements of user experience*
A model for UX encapsulates the key elements that play a role in the formation of UX. The model shows how product features produce an intended product character, consisting of pragmatic and hedonic attributes. This is the designer perspective. When this product is proposed to a user, the user perceives a product character, consisting of pragmatic and hedonic quality. The interaction with the product takes place in the context of the user, resulting in consequences which define how the user feels about the product. This is the user perspective.
- The product character that the designer intends to create is no guarantee for the perceived product character; any user will create a personal version, based on the product's features and the user's personal standards and expectations. The fit to the user's context will result in a judgement of the product's appealingness, emotional consequences, and behavioral consequences. This process will always take place, and the outcome may change with growing knowledge and experience.
- A product's pragmatic attributes are purposed for a product's functionality and usability. This enables the user to fulfill behavioral goals (i.e. do-goals). The perceived functionality and usability in relation to tasks of the user results in a judgement of the product's pragmatic quality.

A product's hedonic attributes are non-pragmatic attributes that address the user's psychological well-being. This enables to fulfill basic human needs (i.e. be-goals). The perceived fulfillment of the user's human needs results in a judgement of the product's hedonic quality.

- Experiencing a product with a certain product character has emotional consequences: satisfaction and pleasure, which are momentary and related to the usage context. Satisfaction is the outcome of the match between expectations and outcomes of the usage. Pleasure is being pleased about a desirable event, and the more unexpected it is, the more intense the pleasure will be. Desirable deviations from the expectations are also pleasing. A product's appealingness depends on the product's ability to trigger positive emotional reactions, combined with perceptions of product attributes and the user's context; an evaluative judgement.

### ➤ *Enabling positive user experience*

Research on positive UXs with technology provides insights on how to enable them.

- For an extensive study, six of the ten basic human needs (autonomy, competence, popularity, relatedness, security, and stimulation) were investigated in relation to affect, attribution, and product perception and evaluation.
    - Most salient needs for positive experiences are relatedness, stimulation, and competence.
    - Need fulfillment is related to positive affect, being a source for pleasure, with stimulation, relatedness, competence, and popularity as most influential needs. It was also concluded that experiences can revolve around a particular need and that different types of experiences can provide particular sets of emotions. Hedonic quality was more strongly related to positive affect compared to pragmatic quality.
    - A direct relation exists between needs and product perception for hedonic quality only, moderated by attribution, meaning that a product's hedonic quality is determined by need fulfillment attributed to the product. It was also concluded that positive affect is a main outcome of need fulfillment, making it a legitimate predictor of hedonic quality.
- While certain human needs are considered more salient and people can have a specific urge for a certain human need, certain needs are not necessarily more important. Fulfilling human needs that are personally important do not make a person happier. Each need should be covered without making this too obvious.
- The relationship between the human needs and question-answering is unclear. There are no examples of earlier attempts to intertwine the two, and examples from other contexts show that human needs are conveyed through an interpretation of the designer to do so. There are no guidelines on how to convey human needs. It is a creative process by the designer followed by testing whether users are indeed triggered to feel a certain way.

### ➤ *Designing user experience*

A simple conceptual model shows three levels (*what*, *how*, and *why*) to consider when designing technology-mediated experiences. The *what* addresses the things people can do through an interactive product (i.e. its functionality), the *how* addresses how to act with the product on an operational, sensory-motor level, and the *why* addresses why people engage in activities: the underlying need(s) for which fulfillment is desired. The *why* is closest to the self (i.e. the user who has internal needs), and that the *how* is closest to world (i.e. the usage context which sets demands for the product). Several other design principles are given for designing a product's UX.

- Understand and address underlying be-goals besides the do-goals that are to be accomplished. Repeat these be-goals during the design process and if needed, introduce particular techniques to fulfill them. Ensure the psychological well-being of the user.
- Consider the context in which users use the product, to understand why they judge a product as appealing, satisfying, and pleasing. Imagine use cases and their contextual factors to evoke realistic experiences.
- Develop representations of the system that is to be developed to get a sense of what the experience might be like, to gather feedback in the form of emotional responses accompanied with explanations.

## 2.3 Main findings and conclusions from the literature review

Main findings and conclusions from the literature review are discussed in this section. The literature review first provided an in-depth review of QASs. A discussion through multiple classifications showed that the technical configurational possibilities are extensive, while the experience of users of such systems is not given much attention. There is no QAS design that consciously considers the UX, and therefore, there is no knowledge on a QAS design addressing UX. This is the research gap. UX is reviewed as well, providing knowledge on how UX unfolds. It enables to approach the design of a system from a UX perspective with the goal of ensuring that users of a system such as a QAS have a positive experience. Enablers of positive UX are found through studies on people's interaction with technology. These theoretical insights on are used to bridge the research gap.

### *Question-answering systems*

To develop a functional QAS, a pipeline must be defined that enables question, document, and answer processing. To guide this, a reflection to the purpose of handling frequently asked customer questions should be made. Certain aspects are determined due to the focus on retrieving answers from FAQ content. This is decisive for the application domain, question type, data source structure, data source characteristics, and answer form. It needs to be specified which techniques are to be used to analyze questions and content and which matching function is to be used to match content with the questions in order to deliver answers. This is highly dependent of the capabilities of Dutch NLP solutions, which are therefore leading. While advanced techniques can be utilized, it is questionable how advanced a QAS covering FAQ content must be. The question similarity technique seems to be a fitting solution for this purpose. However, these findings are purely concerned with the system's functional quality while this research is purposed to enrich the system with UX concepts to enhance the experience that unfolds through its usage. Its importance is emphasized in scarce research towards the experience of QASs, proofing that certain non-functional qualities can better the experience. This leads to the conclusion that, while a full QAS design does require a specification of how to perform question, document, and answer processing, the focus should not be on the technical performance. The focus must be on what can be added to the automation of question-answering to enhance the experience for users.

### *User experience*

A lot of research towards the experience of customers can be found, but only UX specifically addresses how experience unfolds through the usage of a particular system. UX explains the relations between a system (i.e. product features that form a product character) and its user (i.e. perceived product character and the user's context, resulting in certain consequences) that are decisive for what individuals experience when they interact with a system. This provides a foundation to consider system designs from a UX perspective rather than from a purely functional perspective. Studies on people's experiences with technology provided insights on the elements that are most important to achieve positive UX. Especially the perceived fulfillment of human needs (i.e. be-goals) is found to be most important for positive UX. This can be embodied through the system's hedonic quality, for instance by providing closeness and communication between people (i.e. relatedness), providing stimulating insights (i.e. stimulation), and providing opportunities for mastery (i.e. competence). The fact that user perceptions and their context are involved and that particular experiences can revolve around particular needs, emphasize that one-size-fits-all does not hold. The system should ultimately be able to adjust to what is assumed to be fitting for the current user. This leads to the conclusion that a system design should be enriched with hedonic attributes to be able to fulfill human needs (apart from having pragmatic attributes to fulfill behavioral goals), ultimately moderated to the user by taking the user and its context into account.

# 3 EMPIRICAL INVESTIGATION

In addition to the literature review on QASs and UX, an empirical investigation has been conducted for an empirical perspective on both subjects as well. In this chapter, the insights gathered at XBank regarding automated question-answering and experience are discussed, followed by a discussion of observations of QAS implementations of Dutch commercial organizations. Finally, the main findings and conclusions are drawn.

## 3.1 Automated question-answering and experience at ExperienceBank

In this paragraph, relevant insights from XBank are discussed. Internally at XBank, current developments related to automated question-answering are investigated. The lack of a QAS for customers and the internal chatbot are discussed, followed by a discussion of views from multiple XBank employees on experience in the context of automated question-answering.

### 3.1.1 Keyword-based search for customers

XBank is an example of an organization that does not offer a QAS solution to their customers for automated question-answering. They only offer a traditional keyword-based search functionality, that requests the user to use keywords rather than a question to perform a search action. This results in search results (i.e. a list of pages with clickable page headers combined with the first few sentences of the webpage), suggested to contain the desired information. There is no question-answering mechanism. A QAS is purposed to overcome this need of having to manually search for your answer.

### 3.1.2 Chatbot for employees

XBank developed a chatbot for their internal IT service desk, purposed for employees that run into problems. This chatbot is developed in collaboration with vendor Nuance Communications, an American company specialized in speech recognition, natural language, and conversational artificial intelligence. Nuance Communications develops chatbots to answer high-volume/low-intelligent customer questions and judges per question whether interpersonal contact is required (or more suiting). The goal of the chatbot was to reduce the queues for interpersonal service by phone, and it was developed for the internal service desk to test whether the solution could be offered to customers as well. It is advertised on the service webpage of the internal XBank website, accompanied with an image of a female, robot-like person. First trials proved that most employees that used the chatbot prefer it over service by phone (80%). A multichannel specialist at XBank explained that the chatbot works through manually created dialogues, modeled as workflows by specialists in service channels like herself. These dialogues can be triggered by the employee's questions and will guide the conversation to answer the question. To determine which dialogues are purposed for which questions, the chatbot has been trained (supervised training) on historical conversational data (4 months old). When the chatbot does not know which dialogue to pick, an initial dialogue is used to find the most suiting dialogue. A drawback of this approach is the extensive manual work to create these dialogues, and another drawback is that only what has been created can be triggered; readily available content in other formats is not covered. Quite often, the provided answers do not answer the initial question, simply because the chatbot cannot deliver the desired information when a dedicated dialogue is absent. In this case, interpersonal contact is suggested to solve the problem. In the beginning, the chatbot only covered 20% of the service content, which caused a lot of negative reactions and hurt the reputation. While it was emphasized that 80% of the users preferred the chatbot over service by phone by its development team, talks with various employees that experienced the chatbot uncovered that in practice, the chatbot is not that popular because it is often not able to produce helpful answers.

Regarding experience, it is not explicitly addressed in the design of the chatbot. As concluded from research on automated question-answering, the focus is often on providing functionality rather than delivering a certain experience, and the chatbot is no exception. In the case of the chatbot, the goal was to recreate the dialogues that an employee would normally have with an actual person of the service desk. Apart from the dialogue style,

no other human aspects were considered to convey in the chatbot to mimic interpersonal contact. In terms of context, the chatbot only takes into account from which page the employee started the interaction, to get an initial idea of the domain in which an answer was sought after. In terms of measuring the satisfaction of usage, it was questioned whether a suiting answer was provided, how the chatbot was experienced in general, and whether the chatbot or interpersonal contact was preferred. Nuance Communications' sales director for Benelux explained that experience is not something intertwined in their chatbots explicitly. The experience that is normally ensured by CSRs is not translated into the chatbot. The chatbot is a tool to reduce the effort to get to an answer, by enabling people to perform self-service or by getting them in contact with a CSR. The CSR is then responsible for delivering a positive experience. An example brought forward is that a chatbot can recognize when a question concerns the death of a relative. A CSR can help this person out while also showing empathy and understanding, which is assumed to be a more pleasant experience compared to a chatbot conversation.

With being XBank's trial for a 'virtual assistant' solution, it was stated by XBank's lead product manager for online services that in the development of a solution that serves customers, the experience is not a main priority yet. The focus is on functionality because currently, there is no automated question-answering solution at all for customers. The main objective is to reduce the number of CSRs required for delivering customer service, and automation of question-answering (at least for high-volume/low-intelligent questions) is assumed to be contributing to that. And with that objective in mind, experience is something that seems to come afterwards when actual customers can judge the provided solution.

### 3.1.3 Views on experience in automated question-answering

This research has been set up because of the initial question from XBank how a '9+ customer experience' can be realized with the deployment of a digital solution replacing CSRs. This so-called '9+ customer experience' refers to the interpersonal contact context, meaning that every employee should deal empathically with customers, provide them a personalized service, and surprise and amaze them with the service they deliver. 'Personal contact', 'involvement', 'human-centered communication', 'personal support during hard times', and 'thinking with the customer' are examples of values that XBank strives for, but these are not explicitly related to or expressible in the context of automated question-answering. Unstructured interviews with 17 employees of XBank (in the domain of customer service and its IT solutions) provided an idea of what they assume to be of influence on the experience that customers have with an automated question-answering mechanism. Also, two presentations about preliminary findings were given for (in total) 9 business architects of XBank's Distribution Strategy & Architecture, which resulted in insights as well. Table 6 shows an overview of the interviewees' functions. The interviewees are anonymized using interviewee codes. These will be used for referral to the ideas that the interviewees brought forward.

*Table 6. Overview of the interviewees.*

| Interviewee code | Function |
|---|---|
| I1 | Product Manager Digital Customer Service |
| I2 | Product Manager Customer Relationship Management |
| I3 | Business Analyst Internet Banking |
| I4 | Delivery Manager Online Customer Interaction |
| I5 | Lead Product Manager Online Services |
| I6 | Business Analyst Internet Banking |
| I7 | Senior Product Owner Mobile Banking |
| I8 | Multi-Channel Specialist |
| I9 | Digital Product Manager Data, AI & Analytics |
| I10 | Head IT Strategy, Data & Architecture |
| I11 | E-commerce Manager Insurance |
| I12 | Product Manager Digital Customer Service |
| I13 | Business Consultant Customer Insight |
| I14 | Business Manager XBank Customer Service |
| I15 | Business Architect IT Strategy, Data & Architecture |
| I16 | Business Change Manager Cross Channel Service Desk |
| I17 | Senior User Experience Consultant |
| I18 | 9 Business Architects of Distribution Strategy & Architecture |

The insights from the talks and presentations can roughly be divided into (1) functional aspects and (2) non-functional aspects:

1. On the functional side, a much recurring comment was that a good experience with automated question-answering is foremost assumed to be associated with delivering correct answers as fast as possible (I4, I5, I8, I9). This was also expressed as reducing the customer's effort to find desired information (I1, I11, I12, I13). And to be able to do so, the content from which answers are derived needs to be in order, which is therefore indirectly linked to the experience (I2, I5, I6, I7, I18). Subsequently, it was stated that people who start searching for an answer on the website should be able to complete this search, because they deliberately use self-servicing (I11). However, it was also emphasized that the system should not be designed to cover every single question. There will always be cases that require interpersonal contact (e.g. complex questions, personal matters) (I1, I5, I8, I12, I15, I16). For a good experience, it was also suggested to associate asked questions (that cannot be answered automatically) with dedicated CSRs who have knowledge of the topic that is derivable from the question (I6, I12, I14).

2. In the context of the chatbot for customers, experience was explicitly linked to its XBank shell (e.g. the familiar colors, a picture of a person) (I2, I4), which seemed quite blunt. However, in the context of a simple question-answering mechanism, a lot of other non-functional aspects were brought forward. Showing empathy and 'speaking' in a friendly manner was often mentioned as a way to deliver good experience (I3, I5, I10, I11, I14). Considering the context in which a certain type of person uses the system is considered to be of importance (i.e. who uses the system in which kind of situation, and can the system adjust to that) (I2, I6, I8, I11, I14, I17, I18). In relation to that, it was suggested to consider providing relevant information based on the context, person, or question, that the customer might not be aware of (yet) (I12, I14). Setting the right expectations was also a recurring subject (I8, I11, I14, I15, I16), which is related to the fact that people will compare the capabilities to other similar solutions they have encountered (i.e. a subjective opinion); if expectations are rather low while the outcome is (surprisingly) good, the experience is positive. However, if the system performs worse than expected, it will have a negative impact on the experience (I17, I18). Also, if people have bad experiences with similar solutions that they encountered, they might already start off with a negative perception (I6, I17). A suggestion was to create a distinctive appearance to prevent the association with other solutions (e.g. a chat-like interface might cause aversion because most chatbots are rather annoying than helpful) (I6). In line with these comments is the emphasis on trust. Customers should trust the solution, ultimately a higher level of trust that they have in CSRs (I1, I15, I16). Apart from seriousness, fun is also seen as a factor to lift the experience. Fulfilling the question-answer process in a fun way rather than mimicking what CSRs would do during interpersonal contact, might be of value (I2, I10). Lastly, an interesting comment was made about presenting the possibilities for interpersonal contact. Apparently, it is a common strategy to 'hide' this information behind (e.g. placed beneath, after) the information that might enable customers to perform self-service, to prevent lots of phone calls. However, this strategy is actually the opposite of being customer friendly and can cause a negative experience through the created annoyance (I6).

Measuring the experience was mentioned as a way to find out what influences the experience, but it is unclear how to actually measure experience (I4, I8, I14, I15, I16). Currently, measurement scales such as the Customer Effort Score and Customer Satisfaction are used within XBank, but those do not explicitly provide information about what truly is decisive or disastrous for the experience (I13, I14, I18). XBank did however create a typology for their customers, creating four customer types which require a different approach: the self-directed, the validator, the avoider, and the delegator (I14, I17). This typology categorizes customers based on the way they seek information (Chaffey, Ellis-Chadwick, Mayer, & Johnston, 2009). XBank extended this distinction by adding numerous human characteristics and customer features to each type (e.g. being emotional, cooperative, impatient, lively).

## 3.2 Question-answering system implementations of Dutch organizations

To extend the empirical perspective on QASs and their experience aspects, market research is performed through an observation and analysis of websites of various commercial organizations (i.e. organizations that offer products and/or services to people to turn a profit). This, because they are supposed to deliver customer service that accompanies those products and/or services. This is done to identify elements in designs of automated question-answering implementations that enable the handling of frequently asked customer questions. This means that the organization provides a mechanism that analyzes the question as worded by the customer and provides dedicated information as an answer (e.g. an extraction, a webpage) rather than a list of keyword-based search results only.

The requirements for an organization to be included in this explorative observational study is that the organization (1) offers online customer service related to their products and/or services, (2) uses the Dutch language on their website, and (3) has a large customer reach. Based on this explorative research, common design choices can be extracted to define how other Dutch organizations propose a QAS solution to their customers. This, to clarify which design elements other than the question input and answer output are integrated in current Dutch QAS solutions purposed for customers.

In this section, found examples are discussed to provide an idea of current Dutch QAS implementations, followed by a summarization of common elements in the design surrounding the question-answering mechanism. Finally, a QAS design is described using a breakdown structure based on the findings to clarify which elements occur in current QAS solutions.

### 3.2.1 Examples of question-answering system implementations

ING, a.s.r., T-Mobile, Zalando, and KLM are Dutch organizations that have a QAS solution integrated into their website. All five are tried out by proposing reasonable questions (i.e. simple question in the organization's context, assumed to be answered automatically), complex questions (i.e. complex question in the organization's context, assumed not to be answered automatically), and unreasonable questions (i.e. random question, assumed to be out of the organization's content scope). This led to the conclusion that all these organizations' QAS implementation try to answer questions with predefined answers, predefined structures to specify the required information, and dedicated webpages with content (i.e. customer service content webpages). It is likely that the used matching method is question similarity, based on the fact that (it seems that) only predefined answers are returned as responses rather than 'random' extractions of text. This means that all 'answers' are paired with known customer questions in advance to be able to calculate the similarity between the user's question and those predefined questions. The highest similarity score defines which 'answer' should be proposed to the user. This question similarity technique is discussed in 2.1.4. Besides the handling of questions, elements on screen surrounding the question-answering mechanism are observed. Per organization, a brief description of its QAS design is given. Corresponding screen captures are included in Appendix B: Screen captures of question-answering system implementations of Dutch organizations.

***ING***

ING, another Dutch financial institution, has placed the QAS question-bar (i.e. the horizontal bar to type a question in) at their homepage accompanied with a welcoming message, inviting you to find what you're looking for quick and easy. The empty question-bar provides suggestions while typing. The rest of the page is filled with typical homepage elements like site navigation, articles, and common services. The same question-bar is provided at the top on the dedicated customer service webpage. Below, common self-services are presented. Further below, a few FAQ are presented per category, and at the bottom of the page, options for interpersonal contact (i.e. web care, chat, office, phone) are presented. When a question is asked, on a new page, a short answer is provided in a few cases, but the user is always provided with a top three FAQ, followed by numerous search results. See Appendix B: Screen captures of question-answering system implementations of Dutch organizations (ING) for screen captures.

***a.s.r.***

a.s.r., a Dutch insurance company, has a homepage fully dedicated to a QAS solution. The layout is chat-like, and it start with a welcoming message and the invitation to indicate what you're looking for (by choosing a suggested

topic or typing a question). The question-bar tells you "Type je vraag in een paar woorden". A chosen topic or asked question results in options to choose from to navigate to the right piece of information, unless it is clear from the initial question. When a question is not understood, the user is asked to reformulate the question. Since a chat-like layout is used, answers (or options to get to an answer) are presented in the same screen. Answers are presented as shortened texts with referrals to see more extensive information, and three related topics are presented as well including the option for interpersonal contact. See Appendix B: Screen captures of question-answering system implementations of Dutch organizations (a.s.r.) for screen captures.

### T-Mobile

T-Mobile Nederland, the Dutch branch of the German mobile telecommunications company, provides a QAS on their customer service webpage. This page first shows topical questions and FAQ, followed by a question-bar, telling you "Stel zelf je vraag". When typing, suggestions are offered. Below, a roadmap through options is presented which needs to be followed to see which interpersonal contact is available. The answer-screen that the user is redirected to after asking a question shows the question-bar at the top, filled with the initial question. Directly beneath it, the question is repeated, and, in most cases, a short answer is given. Below, search results are presented as well. The proposed answer is accompanied with the question whether the answers is satisfying. If not, a dropdown is shown to indicate what is lacking. See Appendix B: Screen captures of question-answering system implementations of Dutch organizations (T-Mobile) for screen captures.

### Zalando

The Dutch branch of Zalando, a German web shop for shoes and apparel, has a QAS solution implemented on their customer service webpage. On this page, suggested topics are presented first, followed by all help topics. Beneath those, a question-bar is presented inviting you with "Begin een vraag of zoekterm te typen". While typing, suggestions are provided. When a question is submitted, the user is redirected to a specific page with information regarding a single FAQ, regardless of the question being understood. The initial question and the question-bar are absent. Beneath the information, the user is asked whether the answer is helpful. If not, the user can leave textual feedback. At the bottom of the page, options for interpersonal contact are presented. See Appendix B: Screen captures of question-answering system implementations of Dutch organizations (Zalando) for screen captures.

### KLM

KLM, the flag carrier airline of the Netherlands, offers a QAS solution through their homepage (if you click on contact) and their customer service webpage. On both, a question-bar is presented, inviting you with "Vind uw antwoord", and directly beneath it, options for interpersonal contact are offered. On the customer service webpage, this is supplemented with three FAQ and various topics as well. When a question is submitted, the user is transferred to a new page, showing the question-bar at the top filled with the initial question. Below, a short answer is provided, or options are provided to specify the problem. The answer contains a referral to a page with more information if desired. If the question is not understood, the user is asked to reformulate the question. Beneath the answer, related questions are presented. Regardless of whether an answer has been given, at the bottom, the user is asked whether the answer was helpful. See Appendix B: Screen captures of question-answering system implementations of Dutch organizations (KLM) for screen captures.

### More examples

In addition, five more examples of organizations with QAS implementations were found and tried out (i.e. ANWB, Eneco, Essent, McDonalds, and Pathé). They all had a lot in common with the previously discussed examples. Therefore, only the main differentiating elements in these additional examples are briefly discussed. ANWB, Eneco, and Essent explicitly provide direct hyperlinks to self-service actions near the question-bar (in addition to FAQ and generic topics). McDonalds provides suggestions for questions as an introduction and makes a distinction between searching for previously asked questions (i.e. existing question-answer pairs) and asking new questions. For the latter, the user is forced to use Facebook to communicate with a CSR. Pathé uses videos and funny animations (GIFs) in their answers, for example, a video about the order process (when asking about how to buy a ticket) or an animation that shows a scared kid (when asking about a lost item). Also, Pathé adjusts the response when the user's input is not matched to any answer twice, by explicitly suggesting that interpersonal contact might be more suiting.

## 3.2.2   The design of the question-screen and answer-screen

While each QAS implementation differs, common elements and design decisions can be derived from all examples observed. A main distinction is made between the part where the user is supposed to enter a question (referred to as the question-screen) and the part where the system responds to the submitted question (referred to as the answer-screen). The answer-screen is sometimes a dynamically changed question-screen, for example, when a chat-layout is used. Common elements and design decisions in both 'screens' are discussed.

### *Question-screen*
The main element of the question-screen is the recognizable question-bar. This question-bar is generally accompanied with an introducing sentence, stating something about the question-bar's purpose. The question-bar is always presented at the customer service webpage (some organizations decide to place it at the homepage as well for navigational functionality). On the customer service webpage, the question-bar is usually centered at the top of the page. It is generally filled with a sentence hinting that a question can be typed in. Default FAQ and an overview of default (self-)service topics are always presented as well. Their placement is most often beneath the question-bar. When options for interpersonal contact are provided on the customer service webpage (which is not always the case), they are usually positioned at the bottom of the page below all other elements that could enable the user to perform self-service.

### *Answer-screen*
In the answer-screen, the question-bar is always presented at the top, filled with the initial question which can be replaced with a new question. The response is presented below the initial question. Most often, a short textual answer is presented in which (a) hyperlink(s) are provided that redirect the user to more information (i.e. dedicated service content, the usual webpage) or more specific information. When the question is partly understood, the user is usually asked to specify the need through clickable options to arrive at an answer. When the question is not understood at all, the user is always asked to reformulate the question to try again. When an answer is presented, the user is often asked to judge whether the answer was helpful ("yes" or "no"). If the user responds with "no", the user can leave feedback about what is lacking. The provided answer is in most cases not companied with FAQ, (self-) service topics, and/or search results related to the user's question. However, in the dynamic webpages, the initial content is still available (i.e. default FAQ and an overview of default (self-)service topics). The same applies to the interpersonal contact options.

## 3.2.3   The definition of the common practice Dutch online question-answering system

Based on the observed and analyzed examples of QASs in practice provided by Dutch organizations as a self-service channel, the common practice Dutch online QAS can be defined. The used question-answering technique and the structure based on common elements are discussed.

### *Question-answering technique*
As stated, the examples most likely make use of question similarity to produce answers to user questions. This means that the organization has created pairs of questions and answers. The user question can be matched with the predefined questions to determine which question-answer pair is most fitting, leading to the response (as described in 2.1.4). This way, the organization is completely in control of the possible answers which is desired (to prevent 'strange' responses) but it does require manual work. User questions must be defined (e.g. based on actual user questions collected through search, live chat, phone calls) and existing content must be paired to those questions. However, if FAQ content is provided on the website, a lot of these pairs are readily available.

### *Structure based on common elements*
Breakdown structures are created to recapitulate the common elements found in current implementations (see Figure 9). While the implementations vary, a selection of common elements is made based on occurrence, and this defines the common practice for QAS designs in a customer service context. A distinction between three types of elements is made: (1) static elements which are always present, (2) optional elements of which one must be 'picked', and (3) possible elements which are not necessarily present. This latter element type is added because of the equal distribution of the elements being present or absent in the observed examples. Blue boxes refer to elements initially occurring in a question-screen and orange boxes refer to elements occurring in an answer-screen. The order of boxes generally reflects the order of elements.
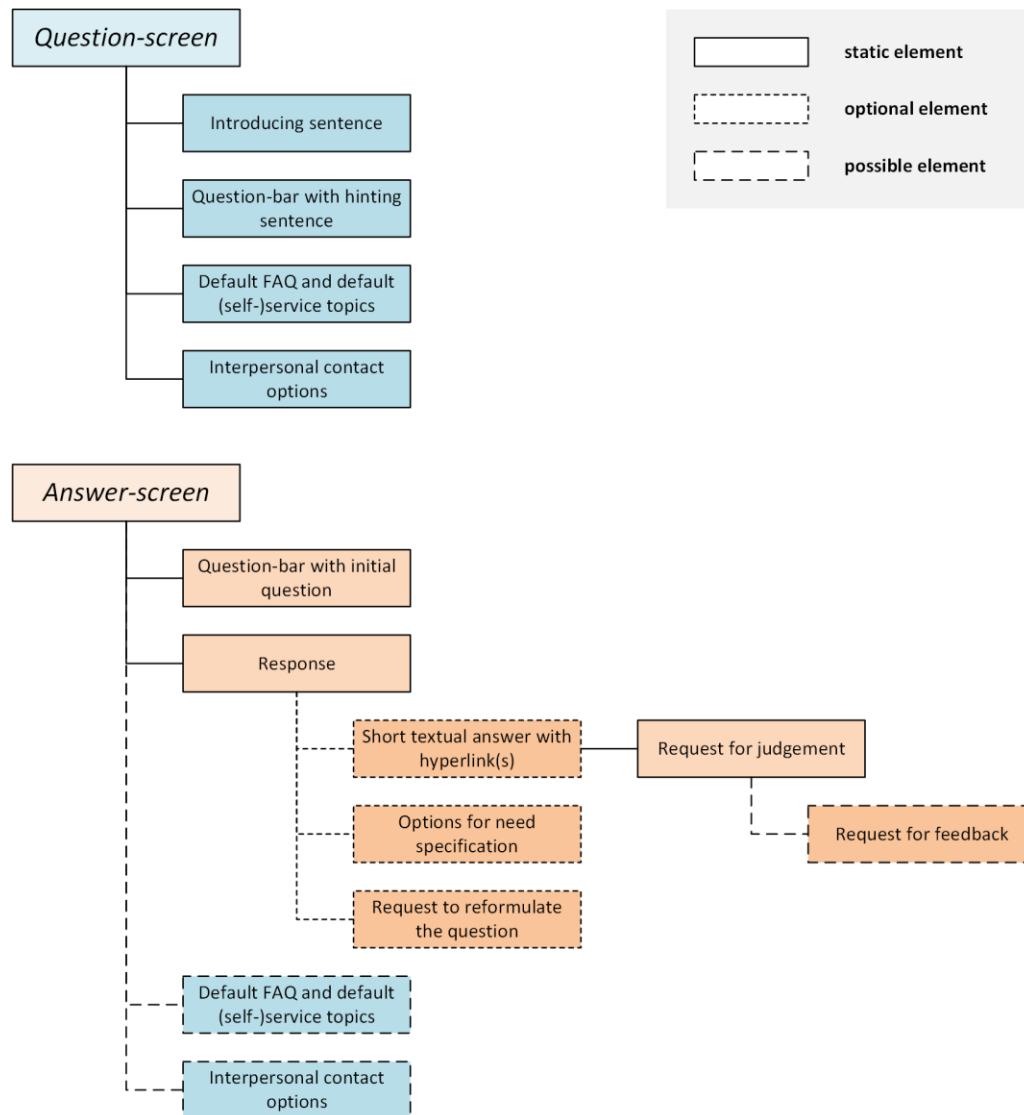
*Figure 9. Breakdown structure of common elements in current Dutch question-answering system implementations.*

In addition to the breakdown structure, visual representations are drawn that represent the element layout of the QAS presented on the website. Figure 10 shows the question-screen on the left and the resulting answer-screen on the right. The answer-screen can be presented as a separate webpage, but it can become part of the question-screen as well (hence the possible elements in blue surrounding the answer-screen). The content of the answer screen is variating based on the response to the initial question (hence the optional response elements). Lastly, the request for feedback is only triggered if the user responds with a negative judgement (hence an optional request for feedback element).
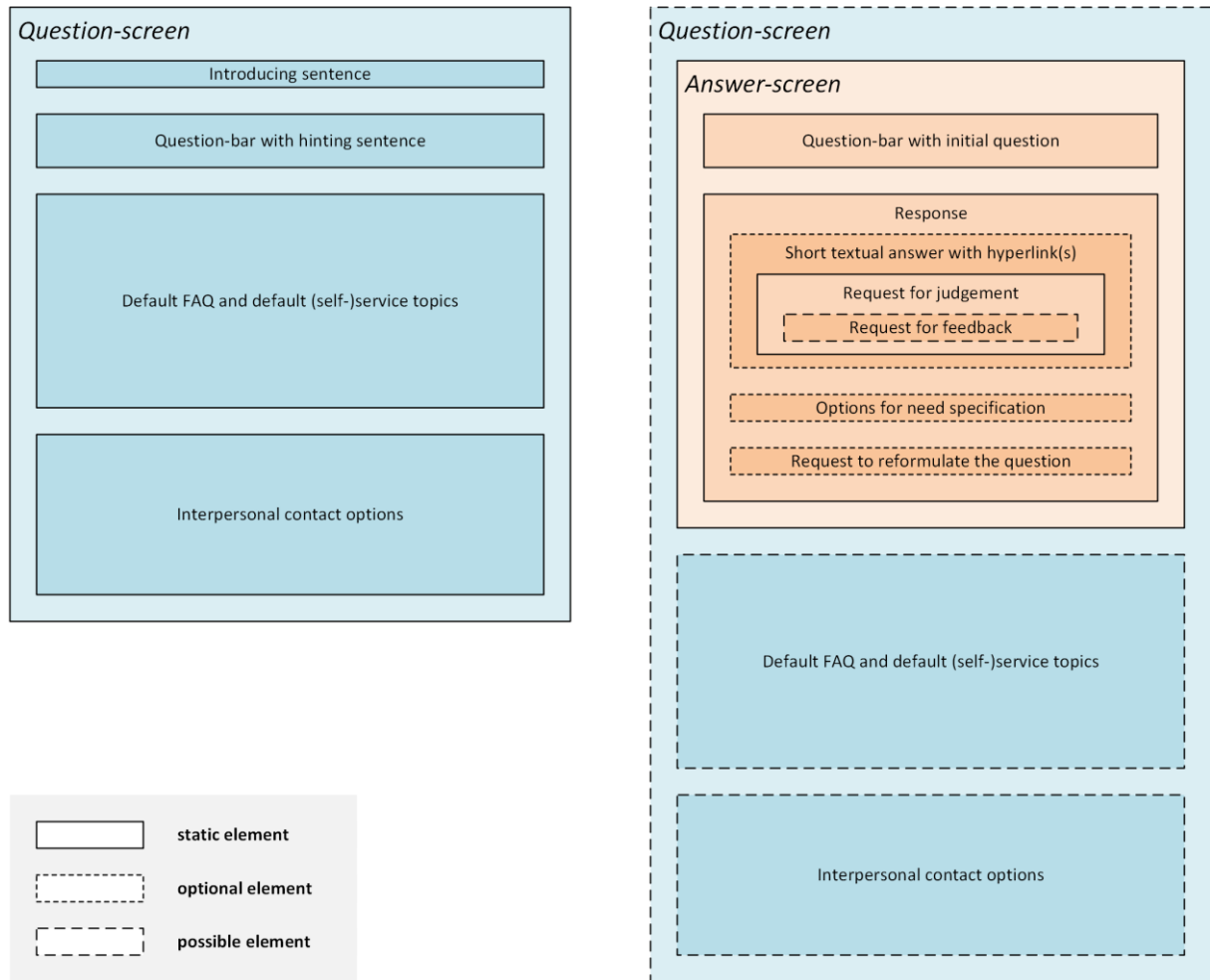
*Figure 10. Element layout representations of the question-screen and answer-screen.*

## 3.3   Main findings and conclusions from the empirical investigation

Main findings and conclusions from the empirical investigation are discussed in this section. The empirical investigation provided insights in XBank's actions and perceptions regarding automated question-answering and the accompanying experience. Currently, there is no automated question-answering solution for customers and while it is in development, the experience of such a solution is not (yet) investigated. However, employees do have opinions about what would influence the experience in this context. Furthermore, based on examples of QAS implementations by various Dutch organizations, certain common elements that are present in the question-screen and/or answer-screen are identified resulting in the definition of the common practice Dutch online QAS. These empirical insights are used to bridge the research gap.

### *Insights on automated question-answering and its experience*

XBank only provides a keyword-based search functionality to their customers on their website. XBank did develop a chatbot meant to solve questions of employees automatically to reduce queues at the internal IT service desk. However, various employees expressed that the chatbot is still not that popular due to not being useful in many cases. Apart from creating dialogues that mimic human conversations, the chatbot is not explicitly enriched from an experience perspective, which was also confirmed by a representative of the chatbot vendor Nuance Communications who stated that experience is not something intertwined in the chatbot. It is rather a tool to enable people to perform self-service, and a CSR will intervene if needed, who ensures a positive experience (e.g. by showing empathy, understanding). XBank is currently focusing on how to create an automated question-answering solution for its customers and the experience of the actual solution will be analyzed and enhanced in a later phase.

Views from multiple employees on experience in automated question-answering resulted in functional aspects and non-functional aspects. Lastly, a way of finding out what influences the experience would be by measuring experience but is unclear how to do this properly since currently used scales (i.e. Customer Effort Score and Customer Satisfaction) do not provide explicit insights on the experience. Based on these insights, it is concluded that experience is not explicitly intertwined in XBank's (internal) chatbot and that the experience is not given attention (yet) in current development. However, in context of automated question-answering, employees expressed aspects assumed to influence the experience. Correct answers, speed, self-service fulfillment, and determined question coverage are indicated functional aspects to positively influence the experience. Showing empathy and 'speaking' in a friendly manner, considering context and person type, setting right usage expectations, propagate trustworthiness, adding a fun factor, and visibly providing interpersonal contact options are non-functional aspects indicated to positively influence the experience as well. Explicitly measuring experience should result in more influence factors.

### *Question-answering system implementations of Dutch organizations*

Examples of QAS solutions that several Dutch organizations (i.e. ING, a.s.r., T-Mobile, Zalando, KLM, ANWB, Eneco, Essent, McDonalds, and Pathé) provide to their customers on their website resulted in a set of common elements and design decisions. These are linked to the 'question-screen' and 'answer-screen' of the QAS solution. This analysis led to the definition of the common practice Dutch online QAS. The used question-answering technique is most likely question similarity to produce answers to user questions, which gives the organization to control the possible answers to prevent 'strange' responses while it does require manual work. Existing FAQ content is however reusable to a certain extent. Common elements found in a question-screen are an introducing sentence, a question-bar with hinting sentence, default FAQ and default (self-)service topics, and interpersonal contact options. For the answer-screen, a question-bar with initial question, a response, and a request for judgement are found elements, which are possibly accompanied by the initial default FAQ and default (self-)service topics and interpersonal contact options. Based on these insights, it is concluded that all solutions most likely make use of the question similarity technique to perform automated question-answering, meaning that more advanced question-answering techniques are not applied (yet) in the defined current common practice Dutch online QAS. Organizations do create their own version of a QAS solution (in terms of question-screen and/or answer-screen layout), but the elements are not that surprising or distinctive, and there is barely anything explicitly experience-related to be found apart from friendly-worded texts.

# 4    RECAPITULATION AND PROSPECT

In this chapter, the main findings and conclusions on QASs and UX from both the literature review and empirical investigation are recapitulated and aligned. The two main subjects are then conceptually combined to theoretically outline what needs to be designed. Based on the insights, it is lastly outlined what is required to work towards a UX-driven QAS and what difference it should achieve compared to current QASs. It prospects what should be conveyed in the to be designed artifact: the *UX enrichment guidelines for QASs*.

## 4.1    Aligning the main findings and conclusions

The literature review revealed that the technical configurational possibilities for a QAS are extensive, while the experience of users of such systems is not given much attention. From an empirical perspective, a similar conclusion was drawn where existing QASs showed a rather simple technique for automated question-answering while also leaving UX largely unaddressed. There are currently no grounds for a UX-driven QAS design, hence a lack of knowledge on how to enrich a QAS from a UX perspective. This is the research gap. UX literature does provide knowledge on how positive UX can be realized for a specific system, and from an empirical perspective, aspects were identified that are in accordance. A combination of the theoretical and empirical insights is used to bridge the knowledge gap. This section recapitulates and aligns the main findings and conclusions from both the literature review and the empirical investigation.

### Question-answering systems

From the literature review on QASs, it was concluded how QASs can be configured on a technical level. It was also concluded that current QAS development is not concerned with ensuring a positive experience. When the experience side of such a system is researched, the focus should not be on the technical performance; what can be added to automated question-answering to enhance the experience for users should be given attention. From the empirical investigation, similar conclusions were drawn. While experience was not explicitly intertwined in XBank's internal chatbot, experience is also not given attention (yet) in current developments for automated question-answering. Examples of QAS implementations of other Dutch organizations are most likely making use of the question similarity technique to perform automated question-answering. Common elements in the systems' structures led to a definition of the common practice Dutch online QAS. The organizations do create their own version of a QAS solution (in terms of question-screen and/or answer-screen layout), but the used elements are however not that surprising or distinctive, and there is barely anything experience-related to be found apart from friendly-worded texts and exceptions found in some examples (e.g. funny animations, explanatory video).

The first research question is thereby answered: *What is a question-answering system and how does it support handling frequently asked customer questions?* From both perspectives it has become clear that automated question-answering functionality can be realized to handle frequently asked customer questions. However, the experience-side of these kind of systems is largely left unattended. The examples of QASs found in literature come across as pallid QASs that, on the surface, basically do not provide anything more than a question-bar to submit questions and answers. However, the techniques used to answer questions can be very advanced, especially compared to the examples of QASs found in practice. The used technique in those examples, question similarity, is rather simple but effective (enough) for answering frequently asked customer questions. Besides, they come across a little less pallid due to being conveyed in the organization's service webpage.

### User experience

From the literature review on UX, it was defined what UX is and it was concluded how UX unfolds when a user uses a system. To create positive experiences, a system design should be enriched with hedonic attributes to be able to ensure the user's psychological well-being through fulfilling human needs apart from having pragmatic attributes to fulfill behavioral goals (i.e. do-goals). Giving the user a feeling of autonomy, competence, popularity, relatedness, security, and stimulation (i.e. fulfillment of be-goals) leads to a by the user perceived hedonic quality, which is found to be most important for positive UX. Anticipations of the user also play an important

role, as well as context. In literature, no direct relations between UX and question-answering were found, and it is assumed that the designer of the system must come up with own ideas. From the empirical investigation, comparable insights were drawn regarding the experience in automated question-answering. XBank employees indicated functional aspects to positively influence the experience. Especially the indicated non-functional aspects assumed to influence the experience (e.g. showing empathy and 'speaking' in a friendly manner, considering context and person type, setting right usage expectations, propagate trustworthiness, a fun factor) are comparable to findings from UX literature. Measuring users' experiences was more heavily emphasized.

The second research question is thereby answered: *What is user experience, what are its concepts, and how does it relate to handling frequently asked customer questions?* From both perspectives it has become clear that there can be more to a QAS than functionality only. UX literature provides clear directions on enriching a system to enhance the experience apart from fulfilling the goal of getting an answer. Considering users, their anticipations, and user contexts, conveying hedonic attributes by addressing human needs (i.e. autonomy, competence, popularity, relatedness, security, and stimulation) in the system's design, and measuring user perceptions are roughly mappable to the aspects derived from the empirical investigation. For example, showing empathy and 'speaking' in a friendly manner can be found in the human need of popularity (feeling liked, respected), trustworthiness can be found in the human need security (feeling safe), and fun can be found in the human need stimulation (feeling enjoyment). Considering context and person type and setting right usage expectations are obvious corresponding with literature findings. A direct relation between UX and question-answering was however absent.

## 4.2  Conceptually combining question-answering systems and user experience

Figure 11 shows a simple conceptualization encapsulating the combination that results in a UX-driven QAS. Knowledge on QASs should be consulted to create a pragmatic system purposed to handle questions and deliver answers (i.e. functionality, symbolically visualized with a QAS architecture), and knowledge on UX should be consulted to create a positive UX (i.e. UX enrichment, symbolically visualized by the six human needs).
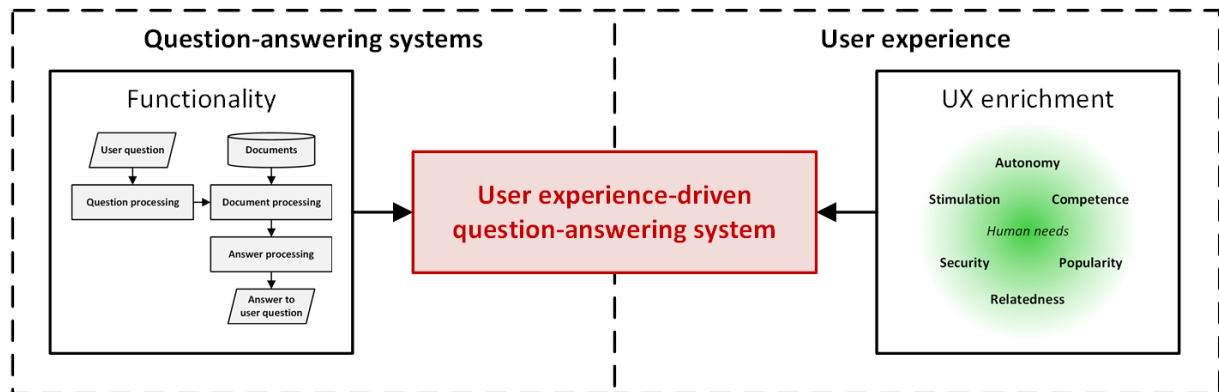


*Figure 11. Conceptual framework visualizing the foundation of a user experience-driven question-answering system.*

An extended conceptual framework is shown in Figure 12. This shows how a UX-driven QAS is positioned between the key elements of UX. The framework is briefly discussed from left to right. A QAS is designed by choosing and combining features for the system based on existing QAS knowledge (white dashed box). To enrich it with UX concepts, instructions regarding elements that enrich a QAS from a UX perspective are required, but such instructions are absent. Therefore, these instructions form the main artifact to be designed and subsequently be validated through expert opinions using an accompanying PoC based on those instructions. This artifact will be referred to as the *user experience enrichment guidelines for question-answering systems* (blue dashed box). The result of designing a QAS using these two sources of knowledge will be two-fold from a designer perspective: (1) the system; a UX-driven QAS with (2) an intended system character consisting of pragmatic and hedonic attributes conveyed by its features. The pragmatic attributes are related to the functionality and the hedonic attributes are related to the UX enrichment conveyed in the design. A user can interact with the system, which is input for a (by the user) perceived system character consisting of a perceived pragmatic and hedonic quality. The combination of this perceived system character and the user's context leads to consequences; the user's

judgements about the system. This outcome conveys the user's feelings towards the system which is summarized as the UX. By adding UX enrichment in the system's design, positive UX is hypothesized to be achievable.
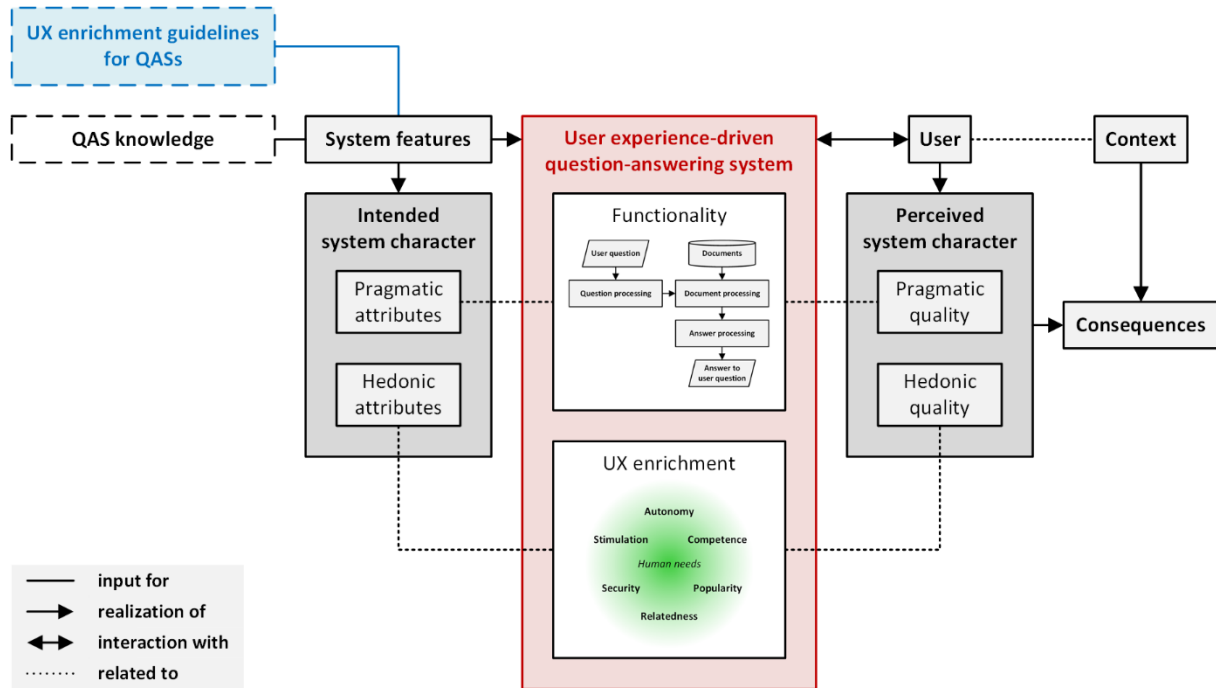


*Figure 12. Conceptual framework addressing key elements of user experience in relation to a user experience-driven question-answering system.*

## 4.3 Towards a user experience-driven question-answering system

Reflecting the definition of UX (as defined in paragraph 2.2.3) to a QAS provides an understanding of the event that is to be enhanced by combining the two knowledge domains. It provides a starting point for outlining the aspects that need to be covered in the UX enrichment guidelines for QASs.

1. The event that is experienced by the user is the user's process of having a certain question, proposing it to the QAS, and retrieving an answer. Proposing a question and retrieving an answer is supported through interacting with the QAS through its user interface.
2. The experience stretches over the duration of the interaction with the QAS:
   - The definite beginning of the experience is the user's initial confrontation with the QAS to start the interaction.
   - The definite end of the experience is the moment the user exits the QAS interaction. While this is assumed to be after receiving an answer to the question (or multiple answers following multiple questions), it could also result in a switch to another customer service channel such as live chat (which could be the case when the QAS cannot deliver an appropriate answer).
3. The interaction with the QAS causes the user to internally process the experience, which is formed by:
   - Judgements, or an evaluative feeling, primarily based on the user's stream of momentary feelings (good-bad) resulting from the interaction with the QAS.
   - Judgements of the interaction with the QAS compared to the user's anticipations of the QAS interaction based on:
     o Expectations of the capabilities of the QAS.
     o Previous experiences with the QAS or similar solutions.
   - Context.

This reflection frames the event (i.e. interacting with a QAS for question-answering) through which the experience of a user unfolds. To work towards a UX-driven QAS, the insights on how positive UX unfolds must be combined with the process of question-answering through a QAS. This, to make sure that the user's internal processing of the experience results in something positive.

Based on the literature review towards QASs, functionality is the only aspect of QASs that has really been given attention. Reflecting this to what is derived from the literature review about UX, it means that QASs currently only supports one particular behavioral goal (i.e. do-goal): asking a question in natural language to retrieve an answer (through an interface). Current QAS designs therefore consist of system features that result in an intended system character consisting of pragmatic attributes rather than a combination of pragmatic and hedonic attributes. This means that users judge a QAS based on its pragmatic quality (i.e. perceived functionality and usability in relation to the do-goal). Figure 13 provides a basic visual interpretation of a QAS that is not enriched with UX concepts, which therefore only supports the explicit do-goal. It starts with a user who encounters the QAS in a certain context. The QAS will try to support the user's do-goal by providing the opportunity to ask a question. The interaction with the QAS will result in an evaluative feeling specific to the user. Eventually, the user is provided an answer (i.e. potential fulfillment of a do-goal) to satisfy the core need. And if the system has no appropriate answer, it should still support in fulfilling the do-goal by offering alternative customer service channels.
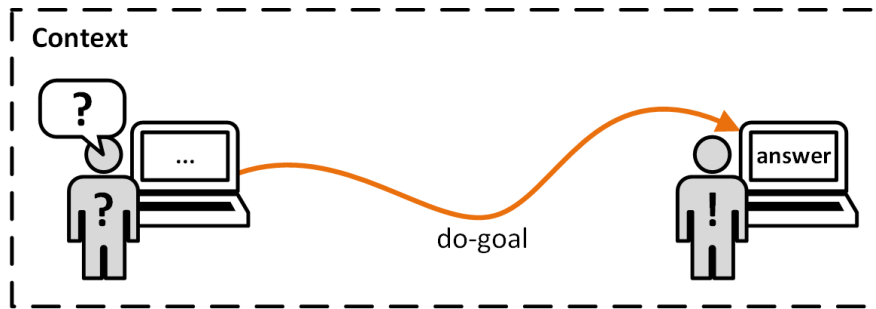


*Figure 13. Interpretation of a question-answering system, only supporting the do-goal.*

### Support the user's be-goals

The lack of hedonic attributes means that the user's psychological well-being is not considered, causing a lack of hedonic quality in the QAS's perceived system character. The consequences resulting from using the QAS are therefore purely a result from the match between the perceived pragmatic quality of the QAS and the user's context. However, based on the literature review towards UX, a system's hedonic quality is more strongly tied to pleasure compared to pragmatic quality, and only the hedonic quality is directly related to need fulfillment and product perception; the subjective side. Therefore, if a QAS supports the goal of retrieving an answer and is consciously enriched with hedonic attributes through addressing basic human needs, it supports users to achieve be-goals as well. Ultimately, this is all moderated to the type of user and its context. It is hypothesized that this results in a positive UX for QAS usage. Figure 14 provides a first basic visual interpretation of a UX-driven QAS, which supports both the explicit do-goal and the more implicit be-goals. Again, it starts with a specific user who encounters the QAS in a certain context. The QAS will support the user's do-goal by providing the opportunity to ask a question, but the QAS will also support be-goals by providing the opportunity to fulfill human needs. This changing the stream of momentary feelings which therefore influences the evaluative feeling specific to the user when interacting with the QAS. Eventually, the user is provided an answer (i.e. potential fulfillment of a do-goal) to satisfy the core need. In comparison to the previous scenario, human needs have now been addressed as well (i.e. potential fulfillment of be-goals) for a positive effect on the user's psychological well-being.
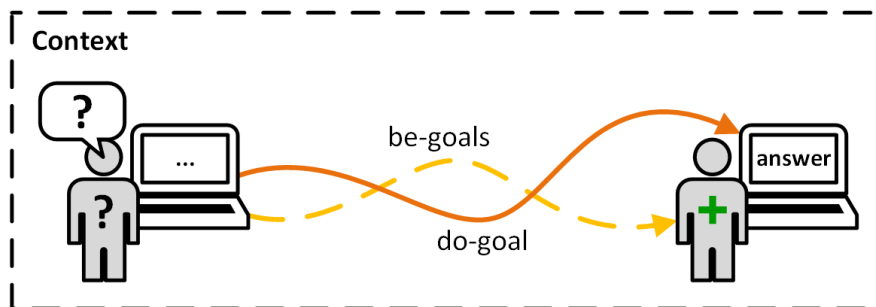


*Figure 14. Interpretation of a user experience-driven question-answering system, supporting the do-goal and be-goals.*

### Ensure that the user has realistic anticipations

Besides the stream of momentary feelings, the anticipations of the user play a role in the internal processing of the experience. Therefore, the capabilities of the QAS should be made clear to influence the expectations the user has of the system. Also, because similar solutions or similar looking solutions unconsciously have already influenced the users' expectations. From the empirical investigation it was concluded that the experience is (surprisingly) good when expectations are low, while a performance worse than expected will be a negative experience. The latter must be prevented. Therefore, realistic anticipations must be ensured from the start of the interaction so that users are aware what they system can and cannot do for them. Figure 15 provides a second basic visual interpretation of a UX-driven QAS, which supports the do-goal and be-goals and ensures that the user has realistic anticipations through clarification (visualized through the speech balloon with an exclamation mark coming from the laptop).
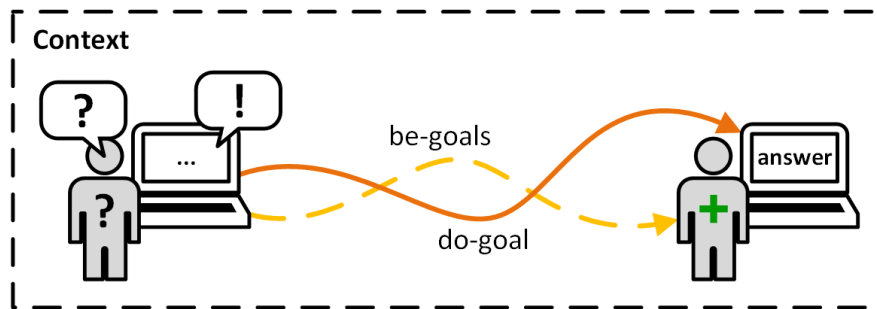


*Figure 15. Interpretation of a user experience-driven question-answering system, supporting the do-goal and be-goals and ensuring realistic anticipations.*

### Provide the organization with insights

While the focus of the design is on sparking positive UX for users, another stakeholder should be addressed: the organization, who provides the QAS to its customers. In the reviewed literature on UX, the organization is only positioned as the designer of a system. Apart from designing a system, an organization should be concerned about the actual consequences rather than only designing a system with assumed consequences based on a theoretical foundation. This means that the design should include a mechanism to collect user feedback to measure what users have experienced. Also, because different types of experience can revolve around a particular need. As stated before, the QAS is ultimately moderated to the type of user and its context. This implies that besides measuring the user opinions on the experience, the type of user and the user context should be identified. Especially this combination of feedback provides the organization valuable insights. It enables the organization to link desired fulfillment of human needs to certain types of users and types of contexts, eventually to evolve the QAS design and ultimately moderate the system to the type of user and its context. Also, because different types of experience can provide a particular set of emotions attached to the context and action. Figure 16 provides a third basic visual interpretation of a UX-driven QAS, which supports the do-goal and be-goals, ensures that the user has realistic anticipations through clarification, and lastly provides the organization with insights based on measured opinions (visualized through the ruler on the right).
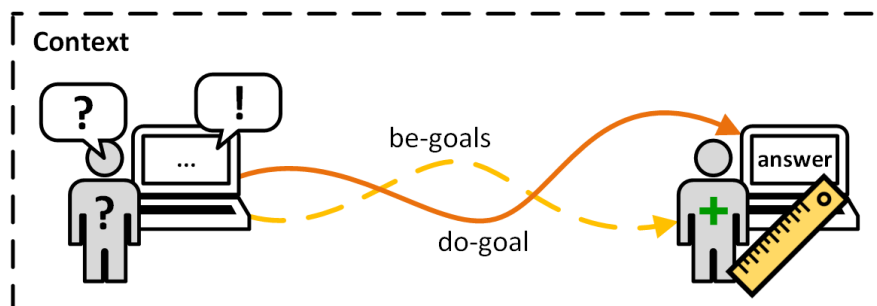


*Figure 16. Interpretation of a user experience-driven question-answering system, supporting the do-goal and be-goals, ensuring realistic anticipations, and providing the organization with insights.*

# PART B: TREATMENT DESIGN

5 THE USER EXPERIENCE ENRICHMENT GUIDELINES FOR QUESTION-ANSWERING SYSTEMS

# 5 THE USER EXPERIENCE ENRICHMENT GUIDELINES FOR QUESTION-ANSWERING SYSTEMS

In this chapter, the design of UX enrichment guidelines for QASs is described. The main goal of these guidelines is to describe how the rather simple mechanism of automated question-answering can be enriched with UX concepts, enabling UX-driven QASs. First, the requirements and context assumptions are described. Next, the full UX enrichment guidelines for QASs. Lastly, the guidelines are applied by means of QAS PoCs: the first PoC represents a generic QAS based on the defined common practice Dutch online QAS. The second PoC represents a generic QAS in which enrichments are conveyed based on the guidelines to exemplify its application. The content of this chapter answers the third research question: *How can user experience concepts be treated in a question-answering system?*

## 5.1 Requirements and context assumptions

In this section, the requirements for the to be designed UX enrichment guidelines for QASs are formulated based on the results of the problem investigation. Furthermore, context assumptions are formulated to outline the context for which the guidelines are designed. This clarifies the assumed context of the QAS and its users.

### 5.1.1 Requirements

The results of the problem investigation led to the conclusion that current QAS designs only address the user's explicit do-goal (i.e. retrieving answers through natural language questions) and do not address any of the more implicit be-goals (i.e. fulfillment of human needs for autonomy, competence, popularity, relatedness, security, and stimulation). UX studies showed that achievement of those be-goals is more strongly tied to positive UX (see paragraph 2.2.5), and interviews with XBank employees uncovered roughly the same areas assumed to be of importance for the experience: empathy and friendliness (e.g. popularity), trust (e.g. security), fun (e.g. stimulation), and being able to complete self-service without CSR interference (e.g. competence) while also visibly providing those alternative interpersonal contact options (e.g. autonomy) (see paragraph 3.1.3). This leads to the conclusion that enriching a QAS with concepts that address be-goals should result in positive UX through interaction with a QAS. From the definition of UX, it was understood that user anticipations also play a role in the unfolding UX because of earlier and similar experiences (see paragraph 2.2.3). Interviews with XBank employees resulted in even more emphasis on user anticipations, suggesting to clarify what can be expected from the QAS to 'control' the user's subjective opinion (see paragraph 3.1.3). From the organization's perspective, it is lastly of interest to get insights from users regarding their experiences to ultimately adjust the QAS to user preferences based on user type and context. XBank employees indicated that it is currently not clear how to truly measure UX, but it is necessary to do so and find out what influences the experience (see paragraph 3.1.3). Based on these insights from the literature review and empirical investigation, the following requirements are formulated to which the design of the UX enrichment guidelines for QASs must adhere (see Table 7).

*Table 7. Requirements for the user experience enrichment guidelines for question-answering systems.*

| Req. | Description |
|------|-------------|
| R1 | The guidelines must instruct what enrichment in a QAS can enable users to start the interaction with realistic anticipations. |
| R2 | The guidelines must instruct what enrichment in a QAS can evoke a feeling of autonomy for users. |
| R3 | The guidelines must instruct what enrichment in a QAS can evoke a feeling of competence for users. |
| R4 | The guidelines must instruct what enrichment in a QAS can evoke a feeling of popularity for users. |
| R5 | The guidelines must instruct what enrichment in a QAS can evoke a feeling of relatedness for users. |
| R6 | The guidelines must instruct what enrichment in a QAS can evoke a feeling of security for users. |
| R7 | The guidelines must instruct what enrichment in a QAS can evoke a feeling of stimulation for users. |
| R8 | The guidelines must instruct what enrichment in a QAS can enable the measurement of user opinions regarding the experience in relation to user types and user contexts. |

In conclusion, the requirements emphasize that the UX enrichment guidelines for QASs require to address the following UX-related concepts: communication about the capabilities (to ensure realistic user expectations), evocation of feelings linked to human needs (to support be-goals), and provision of means to collect a combination of user feedback (to measure the effects of UX enrichments).

## 5.1.2  Context assumptions

The following context assumptions are defined to outline the context for which the UX enrichment guidelines for QASs are designed:
- The QAS is purposed for answering FAQ-type of questions of a commercial organization.
- The QAS can match user questions to predefined questions available in a dedicated database with question-answer pairs, created by the organization.
- The QAS is not able to adjust its presentation to a certain type of person and his/her context.
- The QAS is not a stand-alone tool; it is integrated in a (dedicated) webpage and can be seen as a replacement of the keyword-based search functionality you currently find on organizations' websites.
- The QAS is customer service-oriented with a focus on FAQ-type of questions only and is therefore meant to be presented on a customer service or FAQ page (i.e. not on the organization's homepage).
- The QAS is not integrated in an application; it is a web-only solution, accessible through any device (e.g. computer, tablet, smartphone).
- QAS users are of all sorts, but generally considered to be non-IT experts, often in a situation that is unclear or concerning to them.
- QAS users can be existing customers but also suspects, prospects, or former customers; a log-in cannot be enforced.
- QAS users are aiming at helping themselves before having to reach out for personal help and having interpersonal contact.
- QAS users are familiar with traditional keyword-based searching with a list of pages/topics as a result of the search action, as this is the standard on most websites of commercial organizations, but it is likely that users are familiar with chatbot-solutions as well.

## 5.2 Defining the user experience enrichment guidelines for question-answering systems

In this section, the UX enrichment guidelines for QASs are defined which address the following main areas of attention: (1) managing user anticipations, (2) addressing human needs, and (3) measuring user opinions. The guidelines are formulated including their goal and a short example describing how it could be realized.

### 5.2.1 Managing user anticipations

UX is foremost a result of a person's perceptions and responses that result from the actual usage, but anticipated use is also of influence (ISO 9241-210:2010, 2010). It plays a role in the subjective evaluation of the used system based on expectations of the product and previous experiences with the product or similar products (Hassenzahl, 2003). Basically, anticipated use leads to three possible outcomes: (1) the actual usage is underwhelming compared to the anticipated usage, which is a strong dissatisfier, (2) the actual usage matches the anticipated usage, which is not necessarily good or bad for the experience, or (3) the actual usage surpasses the anticipated usage, which is a strong satisfier. Therefore, it must clearly be communicated to the user what the offered question-answering functionality can and cannot do. This must be conveyed as follows:

| 1.1 | Guideline | State which results can and cannot be expected from the QAS by explaining the covered content areas and the question's level of complexity that the QAS can handle. |
| --- | --- | --- |
| | Goal | Clarifying the QAS's capabilities so that before usage, the user can judge whether the QAS can fulfill his/her information need. |
| | Example | An explanatory text accompanying the question-bar. |

| 1.2 | Guideline | State what the user can submit in the question-bar (i.e. a question in natural language, keywords) to retrieve corresponding information. |
| --- | --- | --- |
| | Goal | Clarifying how to use the system so that it is likely that a user submits effective input. |
| | Example | A placeholder text in the question-bar indicating that a natural language question and keywords can be submitted. |

### 5.2.2 Addressing human needs

The main finding from UX literature is that human need fulfillment (i.e. fulfillment of be-goals) leads to positive experienced affect. While certain human needs were found more salient, it is required to address all human needs (i.e. the set of six human considered to be most important in the context of experiences with technology (Hassenzahl, Diefenbach, & Göritz, 2010)). However, for the context of question-answering, it is not yet clear what the needs represent but it was concluded that designers are supposed to come up with their interpretation. Therefore, based on the descriptions of human needs (Sheldon, Elliot, Kim, & Kasser, 2001), the human needs are 'translated' to the context of a user's question-answering process (see Table 8).

In this paragraph, descriptions of how the human needs can be conveyed in a QAS interaction are given based on both the human needs as defined by Sheldon, Elliot, Kim, & Kasser (2001) and the human needs as defined for the question-answering context.

*Table 8. List of the six human needs and their descriptions based on the context of a customer's question-answering process (in alphabetic order).*

| Human need | The question-answering process must give the user the feeling … |
|---|---|
| Autonomy | … that the process is directed by him/her rather than feeling forced by the organization to use a certain customer service channel in a certain way to reasonably fulfill the process. |
| Competence | … that getting through the process is due to being capable and effective with his/her own actions rather than feeling that his/her competence is put to the test during the process. |
| Popularity | … that they are welcomed as a valued user and that they can provide input about the process which influences their process and that of others to come rather than feeling like just another customer and that his/her input is discarded. |
| Relatedness | … of easily getting in contact with people that care about him/her and his/her questions and found answers rather than feeling secluded and not being able to communicate about questions and/or found answers. |
| Security | … that he/she is in safe hands and is able to prevent escalation of concerning events rather than feeling unsure about a certain situation and being afraid of negative consequences. |
| Stimulation | … of getting enjoyment and pleasure besides getting his/her question answered rather than feeling that asking a question and getting an answer is all there is. |

## *Autonomy*

Autonomy means feeling like you are the cause of your own actions rather than feeling that external forces or pressure are the cause of your action (Sheldon, Elliot, Kim, & Kasser, 2001). In a question-answering context, this is translated to giving the user the feeling that the process is directed by him/her rather than feeling forced by the organization to use a certain customer service channel in a certain way to reasonably fulfill the process. This must be conveyed as follows:

| 2.1 | Guideline | Enable the user to control the process of question-answering (i.e. seeking information) by visibly providing the available options (e.g. the QAS, interpersonal contact options, default FAQ, default (self-)service topics) without oppressing them. |
|---|---|---|
| | Goal | Enabling the user to decide how to seek information without any required actions to uncover new options. |
| | Example | Visibly show the possible (alternative) options as close as possible to the question-bar. |

| 2.2 | Guideline | Only advise the user to consider interpersonal contact if the question clearly does not match any available content. |
|---|---|---|
| | Goal | Enabling the user to decide on follow-up actions. |
| | Example | In the case of an unanswerable question, respond with the advice to try an interpersonal contact option for a follow-up. |

## *Competence*

Competence means feeling that you are very capable and effective in your actions rather than feeling incompetent or ineffective (Sheldon, Elliot, Kim, & Kasser, 2001). In a question-answering context, this is translated to giving the user the feeling that getting through the process is due to being capable and effective with his/her own actions rather than feeling that his/her competence is put to the test during the process. This must be conveyed as follows:

| 3.1 | Guideline | Repeat the user's formulated question to give the impression that that question is being responded to rather than the predefined question only. |
|---|---|---|
| | Goal | Complimenting the user's (formulation of the) question by showing that it caused the response. |
| | Example | Repeat the user's literal input (i.e. submitted question) above the response. |

| 3.2 | Guideline | The process to submit a question and retrieve an answer is easy to fulfill but the user can still be complemented on achieved actions. |
|---|---|---|
| | Goal | Complimenting the user's effectiveness in using the QAS. |
| | Example | Start a response with a celebrating word or message, such as "score!" to confirm a successful search for an answer. |

| 3.3 | *Guideline* | The system must always take the blame in the case of not finding a fulfilling answer. Do not instruct the user to adjust their formulated questions but explain that the system might (still) be unable to help in this particular situation. |
|---|---|---|
| | *Goal* | Emphasizing the QAS's failure to avoid accusations towards the user. |
| | *Example* | In the case of an unanswerable question, respond with an apologizing text and explanation of why the QAS has failed to deliver an answer based on its current capabilities. |

## *Popularity*

Popularity means feeling that you are liked, respected, and have influence over others rather than feeling like a person whose advice or opinion nobody is interested in (Sheldon, Elliot, Kim, & Kasser, 2001). In a question-answering context, this is translated to giving the user the feeling that he/she is a valued user and that he/she can provide input about the process which influences their process and that of others to come rather than feeling like just another customer and that his/her input is discarded. This must be conveyed as follows:

| 4.1 | *Guideline* | Every user must be warmly welcomed as an individual rather than a general visitor. |
|---|---|---|
| | *Goal* | Starting the QAS interaction on a positive, personal note rather than statically. |
| | *Example* | Add a positive, personal word or message in the title, such as "good afternoon!". |

| 4.2 | *Guideline* | Every user must have the opportunity to leave feedback about anything the system conveys (categorizable by content, presentational style, interaction style, and functionality). |
|---|---|---|
| | *Goal* | Enabling the user to express opinions about anything. |
| | *Example* | Showing a button in any screen that takes you to a screen in which the user gets to describe and submit feedback. |

| 4.3 | *Guideline* | It must be emphasized that all 'expert' feedback is more than welcome and will be taken seriously to improve for the user's future use and that of others. |
|---|---|---|
| | *Goal* | Expressing appreciation for the user's feedback who is entitles as usage expert. |
| | *Example* | Invite the user as an expert in the feedback-screen and state that the feedback will help to improve the QAS for him/her and other users. |

| 4.4 | *Guideline* | It must be emphasized that indicating whether the answer was helpful or not helps to improve the QAS. |
|---|---|---|
| | *Goal* | Expressing appreciation for the user's feedback regarding the helpfulness of an answer. |
| | *Example* | Adding a message to the question to indicate whether the answer was helpful or not, stating how the user's feedback helps to improve the QAS. |

## *Relatedness*

Relatedness means feeling that you have regular intimate contact with people who care about you rather than feeling lonely and uncared of (Sheldon, Elliot, Kim, & Kasser, 2001). In a question-answering context, this is translated to giving the user the feeling of easily getting in contact with people that care about him/her and his/her questions and found answers rather than feeling secluded and not being able to easily communicate about questions and/or found answers. This must be conveyed as follows:

| 5.1 | *Guideline* | Always express that there is always a dedicated CSR able to help by showing interpersonal contact options in addition to the QAS's question-bar. |
|---|---|---|
| | *Goal* | Addressing the possibilities for interpersonal contact (i.e. personal assistance). |
| | *Example* | Visually showing the interpersonal contact options with a short explanation how it enables interpersonal contact (e.g. by phone, through live chat). |

| *5.2* | *Guideline* | Users must be able to share findings directly from the page, including the original question, with those who might be involved (e.g. a relative, or in the case of finding an answer for someone else), overcoming the need manually communicating the findings. |
|---|---|---|
| | *Goal* | Supporting the user in communicating about the question and answer with personal close ones in a quick, convenient way. |
| | *Example* | A functionality in addition to the provided answer that enables the user to send an email directly from the answer-screen which automatically contains the asked question and retrieved answer. |

### Security

Security means feeling safe and in control of your life rather than feeling uncertain and threatened by your circumstances (Sheldon, Elliot, Kim, & Kasser, 2001). In a question-answering context, this is translated to giving the user the feeling that he/she is in safe hands and is able to prevent escalation of concerning events rather than feeling unsure about a certain situation and being afraid of negative consequences. This must be conveyed as follows:

| *6.1* | *Guideline* | Emphasize that every question can be answered and that every problem can be solved, possibly already through self-service by using the QAS and otherwise definitely through interpersonal contact with a dedicated CSR. |
|---|---|---|
| | *Goal* | Comforting the user who finds itself in a personally troubling situation. |
| | *Example* | In addition to the question-bar, a comforting message can be added to promise the user that any question or problem will be solved. |

| *6.2* | *Guideline* | Emphasize that the user is at the right place to start solving the problem. |
|---|---|---|
| | *Goal* | Ensuring the user that the arrival at the QAS (i.e. service-page) is a good start. |
| | *Example* | In addition to the question-bar, a message can be added to express that the QAS (i.e. service-page) is the starting point for solving questions and problems are solved. |

### Stimulation

Stimulation means feeling that you get plenty of enjoyment and pleasure rather than feeling bored and under stimulated by life (Sheldon, Elliot, Kim, & Kasser, 2001). In a question-answering context, this is translated to giving the user the feeling of getting enjoyment and pleasure besides getting his/her question answered rather than feeling that asking a question and getting an answer is all there is. This must be conveyed as follows:

| *7.1* | *Guideline* | Show interesting, useful 'fun facts' when encountering the QAS. |
|---|---|---|
| | *Goal* | Adding fun to the rather formal context that might stimulate the user to use the QAS. |
| | *Example* | In addition to the question-bar, it can be shown how many people have (indicated to have) found a helpful answer on the current day or to date. |

| *7.2* | *Guideline* | Besides delivering answers, interesting related articles or related facts can be presented after the presentation of the answer, something to learn from and tell others about. |
|---|---|---|
| | *Goal* | Adding relevant follow-up information in relation to the answer. |
| | *Example* | In addition to an answer, interesting articles can be suggested, such as an article about fraud-prevention after landing on a page about a stolen bankcard. |

## 5.2.3 Measuring user opinions

Lastly, the UX enrichment guidelines for QASs include means to measure user opinions regarding the UX. This is intended as a dedicated tool for the organization to collect systematic insights about the UX of actual users of the QAS. Through feedback from actual users by means of a survey, opinions on the pragmatic quality, hedonic quality, and general product evaluation can be measured. The survey is based on the scales used in the questionnaire used by Hassenzahl, Diefenbach, & Göritz (2010) (i.e. human need fulfillment, experienced affect, product perception, and attribution), but it is slightly simplified and adjusted to the context of a QAS. In addition, the user type is determined, and several contextual factors are identified to be able to relate opinions of the UX to certain groups of users and contexts. Each part of the questionnaire is described. The full questionnaire that it constitutes can be found in Appendix C: User experience questionnaire for question-answering systems.

### Human need fulfillment

Human need fulfillment is measured using a 5-point scale ("not at all" to "extremely") for each human need (i.e. autonomy, competence, popularity, relatedness, security, and stimulation) based on its definition by Sheldon, Elliot, Kim, & Kasser (2001). For example, the user is asked "*Does the system give you the feeling that you are directing the question-answering process?*" This measurement provides insight about the extent to which users think that human needs are conveyed in the QAS's design.

### Experienced affect

Experience affect is measured using the abridged version of the PANAS (Hassenzahl, 2008) to measure the different affective experiences through 10 verbal descriptors for negative affect (i.e. afraid, scared, nervous, upset, distressed) and positive affect (i.e. alert, determined, enthusiastic, excited, inspired). Using a 5-point scale ("not at all" to "extremely") for each verbal descriptor enables the user to express experienced affect. This measurement provides insight about the user's experienced affect by using the QAS.

### Product perception and evaluation

Product perception and evaluation is measured using the abridged version of the AttrakDiff2 questionnaire (Hassenzahl & Monk, 2010). This consists of 4 items to measure pragmatic quality (i.e. confusing-structured, impractical-practical, unpredictable-predictable, and complicated-simple), 4 items to measure hedonic quality (i.e. dull-captivating, tacky-stylish, cheap-premium, and unimaginative-creative), and 2 items to measure general product evaluation (i.e. good-bad and beautiful-ugly). For each item, a 7-point scale is used. This measurement provides insights about the user's perception and evaluation of the QAS.

### Attribution

Attribution is measured with a single question. On a 5-point scale ("very small" to "very large"), users are asked to assess the extent to which the product caused the experience (Hassenzahl, Diefenbach, & Göritz, 2010).

### User type

To determine the user's user type, the typology used to define customer segments based on information need and trust is used (Chaffey, Ellis-Chadwick, Mayer, & Johnston, 2009). This typology defines four types of people seeking information: (1) self-directed, (2) validators, (3) avoiders, and (4) delegators (see Table 9).

*Table 9. Segmentation based on information need and trust. Adapted from Chaffey, Ellis-Chadwick, Mayer, & Johnston (2009).*

| | Ignore advice 'do it yourself' | Depend on advice 'do it with me' |
|---|---|---|
| **Gather detailed information** | *Self-directed* Value: information, control, speed | *Validators* Value: comparison, tables |
| **Don't gather information** | *Avoiders* Value: simplicity, trust, reassurance | *Delegators* Value: online chat, phone support |

To determine what type of user is expressing opinions about the QAS, users are asked whether they generally prefer to gather detailed information or not and whether they generally ignore or depend on advice when they seek information.

### Context

To determine the user's context to a certain extent, four closed questions are asked regarding the social context and physical context. The user is asked whether the QAS was used individually or not, whether people were in the same room/space while using the QAS, whether the QAS was used on a computer, tablet, or smartphone, and whether this was in a comfortable room/space (e.g. at home, at work) or 'on-the-go' (e.g. while walking outside, on the bus).

## 5.3 Applying the user experience enrichment guidelines for question-answering systems

In this section, the application of the UX enrichment guidelines for QASs is demonstrated using generic QAS PoCs. These PoCs are interactive mock-ups that represent two QAS versions: (1) a generic QAS with a structure based on the common practice Dutch online QAS as defined during the empirical investigation, and (2) an enriched version of that same generic QAS, based on the defined guidelines for UX enrichment. The latter is thus a PoC of a generic QAS that contains examples of enrichments as a result from following the UX enrichments guidelines for QASs. Table 10 shows and overview of the mock-ups that have been created for the two generic QAS PoCs.

*Table 10. Overview of the created mock-ups for the generic question-answering system proof of concepts.*

| | | Generic question-answering system proof of concepts | |
|---|---|---|---|
| | | Common practice | Common practice + UX enrichment |
| **Mock-ups** | Question-screen |   Generic question-screen |   Enriched generic question-screen |
| | Answer-screens |   Generic answer-screens |   Enriched generic answer-screens |

The QAS PoCs are characterized as generic because they do not adhere to any type of organization or theme and are therefore neutral in their appearance. For the simulation of interacting with the system, it is chosen to define three possible scenarios: (1) the user submits an answerable question, (2) the user submits an unspecific question, and (3) the user submits an unanswerable question.

The PoCs are created using both Microsoft Visio and Figma. Microsoft Visio is used to create wireframes (i.e. visual representations of the QAS screens) and Figma is used to make those wireframes interactive. In short, Figma makes it possible to link (combinations of) images to the possible screens of, for example, a website. In those screens, areas can be defined that become clickable to trigger an action, such as the appearance of a new screen. Through this, it becomes possible to create a realistic, interactive PoC with mock-ups of the possible screens which can be scrolled through. The Google Chrome browser and a Windows 10 taskbar are also visible for a realistic appearance. Unfortunately, only the visual representations of the screen contents can be presented in this document.

### 5.3.1 Generic question-answering system proof of concept

Based on the defined common practice QAS (see paragraph 3.2.3), the generic QAS PoC is created. Visual representations of the question-screen and three variations of the answer-screen are created (see Appendix D: Screen captures of the generic question-answering system proof of concept). Each screen is briefly discussed.

*Question-screen*
Apart from the default panes at the top and bottom of the screen, the question-screen contains the elements as defined in the structure of a common practice Dutch QAS: an introducing sentence, the question-bar with a hinting sentence, the default FAQ and default (self-)service topics, and lastly the interpersonal contact options. The visual representation of this generic question-screen can be found Appendix D: Screen captures of the generic question-answering system proof of concept (Question-screen).

In the question-screen in the interactive PoC, a mouse click on the magnifying glass shows the three scenarios to choose from to arrive at an answer-screen. Apart from the default panes at the top and bottom of the screen, the answer-screens contain the elements as defined in the structure of a common practice Dutch QAS: the question-bar with the initial question, the response (which defines which of the three answer-screens is shown), and the FAQ and default (self-)service topics and interpersonal contact options that were present in the previous question-screen. The system's response is the variable element and thus different screens are to be presented. Their answer-screen variations are briefly discussed.

### *Answer-screen 1: answerable user question scenario*

The answer-screen in the answerable user question scenario shows a short textual answer with hyperlink(s) (for referral to more information) as the response. Below the answer, the user is asked to indicate whether the answer was helpful or not (by clicking "yes" or "no"). "Yes" leads to a short thank you-message while "no" leads to a textbox with the additional question to explain why it was not helpful. When the feedback is submitted, the thank you-message is shown. The accompanying visual representations can be found in Appendix D: Screen captures of the generic question-answering system proof of concept (Answer-screen 1).

### *Answer-screen 2: unspecific user question scenario*

The answer-screen in the unspecific user question scenario shows a list of most plausible topics (as hyperlinks) to answer the question as the response. Clicking on a topic leads to answer-screen 1. The accompanying visual representation can be found in Appendix D: Screen captures of the generic question-answering system proof of concept (Answer-screen 2).

### *Answer-screen 3: unanswerable user question scenario*

The answer-screen in the unanswerable user question scenario shows a message stating that the question cannot be answered and that the question should be reformulated as the response. The accompanying visual representation can be found in Appendix D: Screen captures of the generic question-answering system proof of concept (Answer-screen 3).

## 5.3.2 Enriched generic question-answering system proof of concept

The presented generic QAS PoC is enriched by following the UX enrichment guidelines for QASs. Visual representations of the enriched question-screen and three variations of the answer-screen are created (see Appendix E: Screen captures of the enriched generic question-answering system proof of concept). Each screen is briefly discussed including referrals to the guidelines presented in section 5.2. The screen visualizations contain colored frames that indicate an enrichment related to a particular topic (i.e. anticipation, autonomy, competence, popularity, relatedness, security, or stimulation). A legend for these colors is shown in Figure 17.
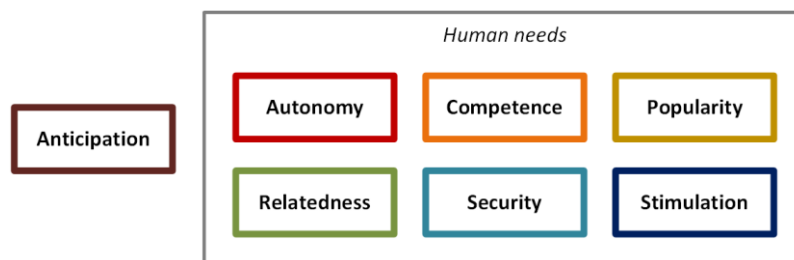


*Figure 17. Legend for the colored frames in the visual representations of the enriched generic question-answering system proof of concept.*

### *Question-screen*

The enriched question-screen consists of the former question-screen elements, adjustments, and enrichments. The title has been adjusted for a more personal welcome following guideline 4.1 (popularity). A subtitle containing some 'fun facts' about the system's usage has been added as suggested in guideline 7.1 (stimulation). The placeholder text in the question-bar is expressing about the possible input formats following guideline 1.2 (anticipation). An explanatory text about the QAS's capabilities is added below the question-bar as suggested in guideline 1.1 (anticipation). Also, a comforting message is added to comfort the user as suggested in guidelines 6.1 and 6.2 (security). Below (and most importantly in the same frame, visible with a single glance), alternative

options are presented: first the interpersonal contact options followed by the default FAQ and default (self-) service topics (in a collapsed manner) as suggested in guideline 2.1 (autonomy). The shown interpersonal contact options are accompanied by a message stating that it is possible to get in contact with a CSR and the options briefly explain how, as suggested in guideline 5.1 (relatedness). Lastly, a feedback-button is placed on the right-side of the screen, inviting the user to express any thoughts about the system as suggested in guideline 4.2 (popularity). A mouse click will trigger a feedback-popup. The visual representation of this enriched generic question-screen can be found in Appendix E: Screen captures of the enriched generic question-answering system proof of concept (Question-screen).

Again, in the question-screen in the interactive PoC, a mouse click on the magnifying glass shows the three scenarios to choose from to arrive at an answer-screen. The enriched answer-screens consist of the former answer-screen elements, adjustments, and enrichments. Again, the system's response is the variable element and thus different screens are to be presented. At the bottom of each answer-screen, the former question-bar is presented, accompanied by the 'fun facts'-subtitle (guideline 7.1, stimulation) and with the hinting placeholder text (guideline 1.2, anticipation). The default FAQ and default (self-)service topics (in a collapsed manner) are presented at the bottom. This, due to following guideline 2.1 (autonomy). In each screen, the former feedback-button is placed on the right-side of the screen, inviting the user to express thoughts about the system as suggested in guideline 4.2 (popularity). The enriched answer-screen variations are discussed.

### *Answer-screen 1: answerable user question scenario*

Following guidelines 2.1 (autonomy) and 5.1 (relatedness), the enriched answer-screen in the answerable user question scenario presents the interpersonal contact options at the top (in collapsed manner), expressing that it is possible to get in contact with a CSR. Below, the response to the question is presented. First, the user's question is repeated, and a message is added to celebrate the successful action as suggested in guidelines 3.1 and 3.2 (competence). Below the answer, a share-button is added accompanied by a text that the answer can be shared as suggested in guideline 5.2 (relatedness): a mouse click will trigger a share-popup, in which the user can provide personal information and a personal message. A mouse click on the share-button in this popup will trigger an email to be send containing the user's name, the user's question, the answer, and the personal message if added. A confirmation message will be shown in the share-popup which can be closed to return to the answer-screen. In this screen, below the share-button the user is asked to indicate the helpfulness of the answer (by clicking "yes" or "no"), and it is stated how this feedback helps to improve as suggested in guideline 4.4 (popularity). "Yes" leads to a thank you-message, expressing that it is nice that the user managed to find a helpful answer as suggested in guideline 3.2 (competence). "No" leads to an apologizing message as suggested in guideline 3.3 (competence). The user is asked to indicate why it was not helpful and multiple-choice options are added for convenience while it is still possible to add an explanation. When the feedback is submitted, the thank you-message is shown stating that it helps to improve for the user and fellow users as suggested in guideline 4.3 (popularity). Furthermore, a link to an article is presented of which the content relates to the answer as suggested in guideline 7.2 (stimulation). The accompanying visual representations can be found Appendix E: Screen captures of the enriched generic question-answering system proof of concept (Answer-screen 1).

### *Answer-screen 2: unspecific user question scenario*

Again, by following guidelines 2.1 (autonomy) and 5.1 (relatedness), the enriched answer-screen in the unspecific user question scenario presents the interpersonal contact options at the top (in collapsed manner), expressing that it is possible to get in contact with a CSR. Below, the response to the question is presented. First, the user's question is repeated as suggested in guideline 3.1 (competence) and a list of most plausible topics (as hyperlinks) to answer the question is presented as the response. Clicking on a topic leads to answer-screen 1. The accompanying visual representation can be found in Appendix E: Screen captures of the enriched generic question-answering system proof of concept (Answer-screen 2).

### *Answer-screen 3: unanswerable user question scenario*

The answer-screen in the unanswerable user question scenario first repeats the user's question as suggested in guideline 3.1 (competence), followed by a message stating that the system failed to answer the question and an explanation of why this happened as suggested in guideline 3.3 (competence). Also, a comforting message is added to comfort the user as suggested in guidelines 6.1 and 6.2 (security). The interpersonal contact options are presented directly beneath the response (guideline 2.1, autonomy, and guideline 5.1, relatedness). The accompanying visual representation can be found in Appendix E: Screen captures of the enriched generic question-answering system proof of concept (Answer-screen 3).

### *Feedback-popup*

The feedback-popup that shows when the feedback-button is clicked addresses the user as expert and starts with a message to express appreciation for any feedback as suggested in guideline 4.3 (popularity). Multiple-choice options are added for convenience of indicating the related feature(s) as suggested in guideline 4.2 (popularity), and an explanation can be added next. A mouse click on the send-button in this popup leads to a confirming thank you-message stating that it helps to improve for the user and fellow users as suggested in guideline 4.3 (popularity). The popup can be closed to return to the former screen. The accompanying visual representations can be found in Appendix E: Screen captures of the enriched generic question-answering system proof of concept (Feedback-popup).

### *User experience questionnaire-popup*

While all the discussed screens and popups are to be triggered by the user, the system must approach users for the UX-questionnaire based on their activity (e.g. after three questions are asked and answers have been found). A UX questionnaire-popup must appear, addressing the user as expert because of the multiple usages and expressing appreciation for feedback as suggested in guideline 4.3 (popularity). The user is asked to fill in a UX questionnaire. "Yes" leads to a new page on which the UX questionnaire for QASs can be filled in. The accompanying visual representation can be found in Appendix E: Screen captures of the enriched generic question-answering system proof of concept (User experience questionnaire-popup).

# PART C: TREATMENT VALIDATION

# 6 VALIDATING THE USER EXPERIENCE ENRICHMENT GUIDELINES FOR QUESTION-ANSWERING SYSTEMS

The UX enrichment guidelines for QASs are validated by expert opinions. In this chapter, it is described how this leads to qualitative data to conclude on the designed guidelines. First, the used validation method is discussed. Thereafter, the characteristics of the validation procedure are discussed.

## 6.1 Validation method

In this section, the validation method is discussed. An explanation of the validation by expert opinions is given, the objectives of the validation are described, and the selected experts for the validation are discussed.

### 6.1.1 Validation by expert opinions

The UX enrichment guidelines are validated by expert opinions (Wieringa, 2014) to gather qualitative data. The guidelines are presented to multiple experts who then imagine to what extent they can support in enriching a QAS in terms of its UX. Predictions of the effects and thus points of adequacy and shortcomings will be the result. The experts' imaginations of the guidelines in the QAS context and their elaboration on the predicted effects serves as the validation model which is observed. Because the imagination of experts is of importance, the guidelines and foremost its application should be explained in a realistic problem context for reliable opinions. Experts must also be asked to elaborate their opinions. Indications for improvement opportunities for the guidelines are logically foremost a result of negative opinions and those are therefore more useful than positive opinions.

### 6.1.2 Validation objectives

The main objective of the validation is to gather expert opinions on the UX enrichment guidelines for QASs, but this objective is achieved using two separate objectives:
1. Gathering expert opinions on how to enrich a QAS with knowledge of the consulted UX theory and a demonstration of the common practice QAS PoC, but without provision of the actual guidelines.
2. Gathering expert opinions on the designed guidelines based on knowledge of the actual guidelines and a demonstration of the enriched common practice QAS PoC for its exemplified application.

The first objective is meant to find out to what extent the experts would come up with similar ideas for enrichment from their perspective as a professional, which provides unbiased opinions on possible enrichments. The second objective is the actual validation of what has been designed, in which they judge the particular (guidelines for) enrichments. The validation procedure is based on fulfilling these two objectives and therefore, it is possible to speak of two phases during the validation procedure.

### 6.1.3 Experts

The experts for the validation were selected within XBank based on their knowledge of UX and their familiarity with XBank's customer service domain. Multiple people within XBank have been approached with an explanation of the research purpose and with the question to participate in the validation procedure and/or provide names of colleagues who might be able to participate. Six experts with different backgrounds in terms of their function were selected and invited. Individual validation sessions of one hour were scheduled with them, and eventually only one expert could not make it. Therefore, a total of five validation sessions were conducted and thus the goal for the number of experts was reached. Table 11 shows an overview of the experts' functions and years of employment at XBank that indicate why they are suitable candidates. As shown, all the experts have different functions, which provides a broader picture during the validation due to multiple perspectives. The experts are anonymized using expert codes. These are used for referral to experts in the discussion of the results.

*Table 11. Characteristics of the participating experts.*

| Expert code | Function | Employed at XBank for |
|---|---|---|
| E1 | Senior User Experience Researcher | 20 years |
| E2 | Product Manager Customer Experience & Channels | 6 years |
| E3 | Lead Customer Experience Designer | 21 years |
| E4 | User Experience Consultant | 25 years |
| E5 | Lead User Experience Designer | 11 years |

## 6.2 Validation procedure

In this section, the executed validation procedure is defined. First, the setting of the validation sessions is described. Thereafter, the artifacts that were created to use in the validation sessions are described. Lastly, the process of a validation session is described using a visualization of the process.

### 6.2.1 Validation setting

The validation sessions were organized in such a way that experts were at ease and external influencing factors were minimized. The validation sessions were organized at the main office of XBank, which is where all the participating experts are mainly based. The same quiet, closed-off room that has room for approximately eight people was used for each session to minimize the chances of disturbance and distractions. Before starting the session, the expert was offered something to drink. At the start of the session, the expert was asked permission to make an audio recording of the session to be able to process the discussed contents afterwards. Every expert gave permission. None of the sessions were interrupted or abruptly ended. The one-hour timeframe was just enough to complete each session in a calm manner. In conclusion, all sessions were executed as desired.

### 6.2.2 Artifacts used in the validation sessions

Several artifacts were prepared for the validation sessions: two presentations, two QAS PoCs, and a handout.

#### *Presentations*
Two presentations were prepared for the validation sessions. The first presentation was purposed for a brief explanation of the research background. This covered an explanation of a QAS, the motives for the research, the research goal, the understanding of UX as defined in the research, and an outline of the topics covered in the designed UX enrichment guidelines for QASs. The second presentation was purposed for discussing the designed UX enrichment guidelines for QASs, and therefore covered all guidelines in full.

#### *Proof of concepts: ExperienceBank question-answering system*
For the validation by expert opinions, it is required that the experts can imagine the application of the designed guidelines for realistic predictions on its effects. Therefore, two additional QAS PoCs have been realized for the validation sessions. The QAS PoCs for the validation adhere to the organizational context of XBank to simulate as if XBank has a QAS. These two PoCs represent the exact same structure of the two generic QAS PoCs, but the neutral character is replaced by XBank's character in terms of content and presentation. For example, the three scenarios are now based on three actual user questions, which reflect the answerable, unspecific, and unanswerable question. Furthermore, elements such as interpersonal contact options are now the actual options that XBank offers. Basically, all neutral elements are replaced by how XBank would convey them. In the enriched QAS PoC, the enrichments are indicated with large colored arrows instead of colored frames because those colored frames were less clear in this PoC. For now, using these XBank QAS PoCs is the best attempt at providing an imaginable 'problem context' for the experts in which they can predict the effects of the designed 'artifact'. For confidentiality reasons, visual representations of the XBank QAS PoCs are not provided in this document.

#### *Handout: user experience enrichment guidelines for question-answering systems*
One handout is used in the validation sessions. This handout is a printed version of the UX enrichment guidelines for QASs that is given to the expert for convenience. This, because the guidelines are quite extensive, yet each guideline is only shown for a short amount of time during the presentation. The handout also contains the color legend (as shown in Figure 17) for clarification of the colored arrows shown in the enriched XBank QAS PoC.

## 6.2.3 Process of the validation sessions

The process of the validation sessions is simple but effective. A visualization of the process is presented in Figure 18. The process starts with the researcher and expert coming together, and the first activity is the presentation of the research background followed by a demonstration of the XBank QAS PoC. Obviously, the first two artifacts are used: (1) the research background presentation and (2) the XBank QAS PoC. After finishing this activity, the researcher asks the expert to give opinions on how the XBank QAS can be enriched. If required, the researcher asks follow-up questions to hear more opinions or elaborations. This activity results in expert opinions on how to enrich the XBank QAS. With this, the first phase of the session comes to an end. The next phase starts with the activity which consists of the researcher presenting the UX enrichment guidelines for QASs and demonstrating how its application enriched the XBank QAS PoC. The expert is also handed a printed version of the UX enrichment guidelines for QASs. Obviously, the last three artifacts are used here: (3) the UX enrichment guidelines for QASs presentation, (4) the enriched XBank QAS PoC, and (5) the UX enrichment guidelines for QASs handout. After finishing this activity, the researcher asks the expert to give opinions on the guidelines and the enriched XBank QAS. If required, the researcher asks follow-up questions to hear more opinions or elaborations. This activity results in expert opinions on the UX enrichment guidelines for QASs. With this, the second phase of the session has come to an end. The expert is then done with the session. The last activity is for the researcher, who must process the gathered expert opinions to have results per phase that can be analyzed in conjunction with the results from the other validation sessions. The session then comes to an end.
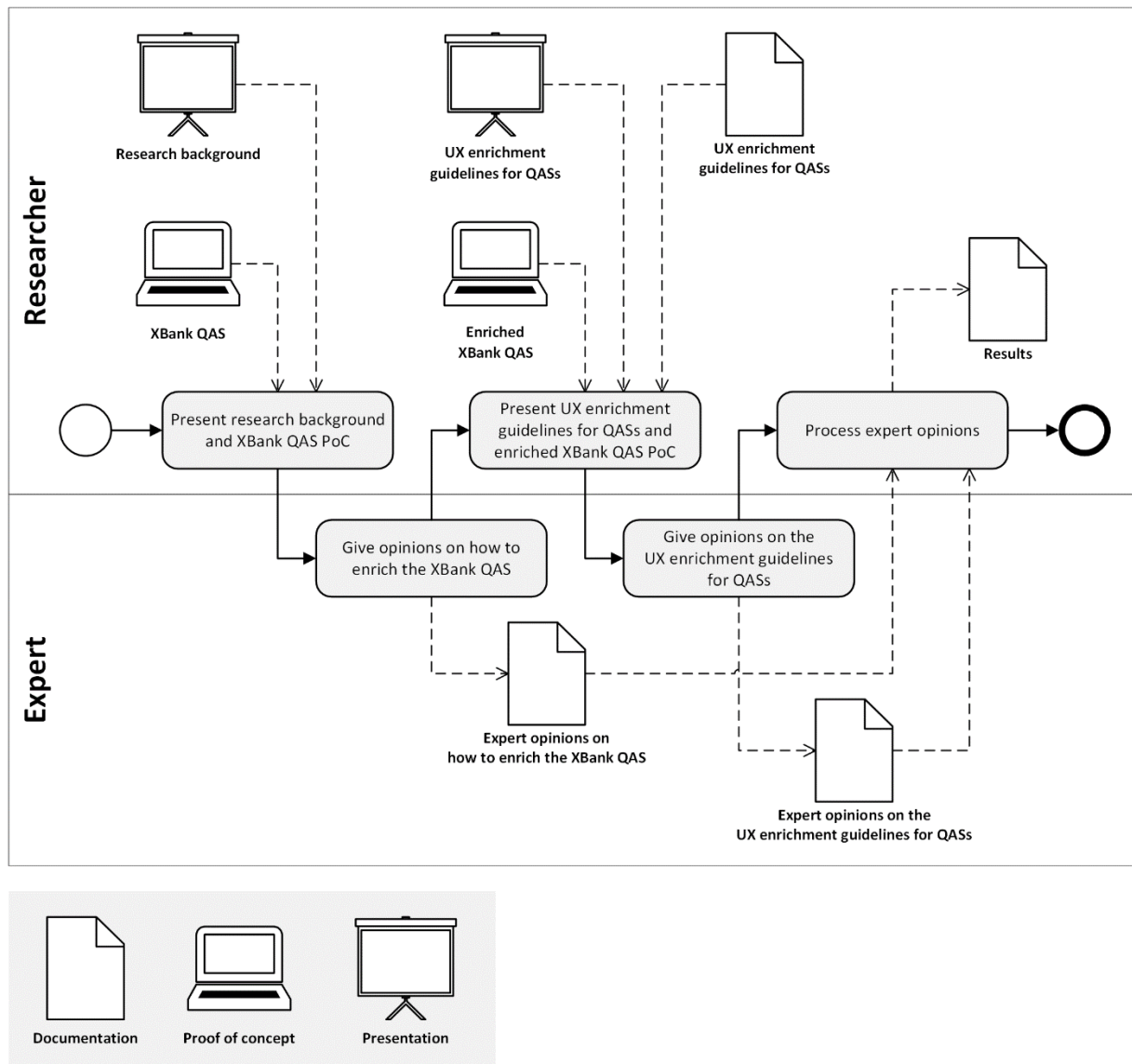


*Figure 18. Process of the validation sessions.*

# 7  RESULTS

In this chapter, the results from executing the validation procedure are presented and discussed. First, the gathered qualitative data is described to provide an overview of the data collection and its characteristics. Next, all results are presented, with a main distinction between the two phases in the validation sessions. In phase one, the experts explained their ideas for UX enrichment based on an understanding of the research context and the first XBank QAS PoC. In phase two, the experts gave their opinions on the UX enrichment guidelines for QASs based on an understanding of the guidelines and its application in the second XBank QAS PoC.

## 7.1  Description of the data

The validation procedure resulted in five audio recordings, each with a duration between 45 and 55 minutes (4 hours and 16 minutes in total). These recordings contain the discussions with the experts, withholding qualitative data. This data is processed using NVivo 12, a tool to store, organize, categorize, and analyze qualitative data (QSR International, 2019). The data can mainly be divided into two parts and can be summarized in numbers using the node tree used to code the comments. A comment here refers to a statement expressed by an expert which contains meaningful information. This can be a short statement of several words, but also a full argument of multiple sentences. When an expert only nodded or verbally agreed with the researcher's statements (by saying "yes" only), it was not registered as a comment. The two parts refer to the two phases during the validation sessions. The first phase resulted in 61 coded comments from the experts based on the research background and the first XBank QAS PoC (based on common practice). The second phase resulted in 84 coded comments from the experts based on the explained UX enrichment guidelines for QASs and the enriched XBank QAS PoC. The topics that form the node tree used for coding are derived from the structure of the guidelines. In addition, some comments were not directly relatable to the guidelines and therefore registered as other remarks. These comments were either focusing on additional satisfiers and dissatisfiers or pragmatic improvements (i.e. related to functionality or usability). Also, some comments were classified as being concluding (i.e. in general, overarching multiple topics). Table 12 and Table 13 present the number of comments per topic and the number of experts that commented on the topic.

*Table 12. Summarization of the comments from the first phase of the validation procedure.*

| Topic | | Comments | From # of experts |
|---|---|---|---|
| Managing user anticipations | | 7 | 5 |
| Addressing human needs | Autonomy | 7 | 5 |
| | Competence | 8 | 5 |
| | Popularity | 3 | 3 |
| | Relatedness | 2 | 2 |
| | Security | 4 | 4 |
| | Stimulation | 6 | 3 |
| Measuring user opinions | | 0 | 0 |
| Other remarks | Satisfiers | 9 | 3 |
| | Dissatisfiers | 4 | 3 |
| | Pragmatic improvements | 9 | 3 |
| Concluding opinions | | 2 | 2 |

Based on the number of comments collected in the first phase, it can be assumed that managing the user's anticipations, and addressing autonomy and competence are the more important aspects since these topics were commented on the most and by all experts as well. Apparently, relatedness was not brought forward much. Also, it is apparent that the experts came with many other remarks, indicating that there are possibly points of attention regarding the UX that are not covered by the topics included in the guidelines. Remarkable is the fact that no comments were given on measuring user opinions, indicating that experts might initially not consider it part of realizing UX enrichment for QASs.

*Table 13. Summarization of the comments from the second phase of the validation procedure.*

| Topic | | Comments | From # of experts |
|---|---|---|---|
| Managing user anticipations | | 10 | 4 |
| Addressing human needs | Autonomy | 5 | 4 |
| | Competence | 6 | 3 |
| | Popularity | 8 | 4 |
| | Relatedness | 2 | 2 |
| | Security | 3 | 3 |
| | Stimulation | 6 | 4 |
| Measuring user opinions | | 12 | 4 |
| Other remarks | Satisfiers | 1 | 1 |
| | Dissatisfiers | 2 | 1 |
| | Pragmatic improvements | 7 | 3 |
| Concluding opinions | | 22 | 5 |

Based on the number of comments collected in the second phase, it can be assumed that managing the users' anticipations and measuring the users' opinions are controversial topics with significantly more comments compared to the (individual) human needs. The human needs are somewhat evenly distributed with popularity being most discussed. Relatedness is again the least discussed. In comparison to the first phase, fewer other remarks were brought forward which is probably related to the fact that the guideline topics were clearer for the experts at this point. Many concluding comments were provided which seems logical since a more overall opinion can be given about the UX enrichment at this point.

## 7.2 Expert opinions on how to enrich a common practice question-answering system

In the first phase, the experts explained their ideas for UX enrichment based on an understanding of the research context and the first XBank QAS PoC. The results from this first phase are presented in this section. The paragraphs are based on the node tree used for coding.

### 7.2.1 Managing user anticipations

The experts emphasized the importance of managing user anticipations. Based on the comments by E2 and E4, it is required to explain what content is covered by the system and which content is not. E4 stated to do that "so that you know what can be answered and what not, to judge whether you want to use the system". Also, letting the user know that natural language can be used as well as keywords is argued to be important by E3. E5 adds to that by stating that a search-bar might incline the user to use it as a traditional search-functionality, while a chat appearance more likely tends the user to really ask a question in natural language. Lastly, E4 expressed that from doing research with users, she learned that users have a need for clarity about a system being either an autonomous system or a way of communicating with a real person.

### 7.2.2 Addressing human needs

This paragraph discusses the experts' comments on how to enrich a QAS in relation to the human needs. Some comments were explicitly linked to these human needs by the experts but there were also comments that rather unintentionally touched upon these topics.

#### *Autonomy*
E1 argues that "a feeling of autonomy is naturally with this type of system because the user does everything himself". E4 argues that a feeling of control over the process is important. The user must be aware of the possibility of quickly finding an answer with the system but must also know the options for interpersonal contact. E5 states that it is important to prevent the user to find a dead end; interpersonal contact options should always be available and nearby "because nothing is as frustrating as not finding your answer, and if you don't succeed you still want to be helped quickly" (E5). Also, E2 argues that users should never be commanded to reformulate their question; "either choose to reformulate or get in contact with a CSR". E4 states that a user should also never be commanded to get in contact; "they will do so only if they want to".

#### *Competence*
Most comments were made regarding a feeling of competence. A few are directed at the response when the system fails to deliver an answer. "It is annoying to hear that you need to reformulate your questions", says E2. E3 explained that the response of the system can be nicer and that pointing at the user is very bad for the user's feelings towards the system; "it is not a good error message". E5 also stated that the user should never be given the feeling that he did something wrong. E1 adds to it by stating that it should be explained to the user why the system failed. E1 and E4 both state that it can be confirmed or complimented that the user did something good if an answer was found. However, it should be subtle, otherwise the user will feel treated childlike (E4).

#### *Popularity*
The scarce comments regarding a feeling of popularity address that the possibility of providing feedback can be utilized, however, "it shouldn't be in-your-face, and only if you know that the user has its answer" (E3). E5 adds that it is possible to express that you understand the user and show appreciation by, for example, expressing that it is nice to see the user making use of the system for asking his question.

#### *Relatedness*
The least touched upon human need is relatedness. E5 however expressed that it is an important aspect in terms of relatedness between the user (i.e. customer) and the organization; "he must feel special; he did not randomly ask XBank a question". And if the system fails to enable the user to fulfill self-service, it must be emphasized that XBank is interested in starting an interpersonal conversation. E3 adds to that by stating that for certain types of questions, interpersonal contact is simply the only suitable option.

### *Security*

E1 argued that a feeling of security does not really apply in the QAS context and E3 only stated that it is not represented in the common practice QAS. E5 linked a feeling of insecurity to when users ask an unanswerable question and are implicitly blamed for it; "this is the last thing you want; users might get an aversion towards the system". E4 did touch upon security more elaborately; "it is good to comfort a user, especially in panicky situations because the human brain does not function appropriately anymore". E4 expressed that answers for such situations, that can be time-critical, should be prepared differently by providing comfort first, followed by an explanation of required actions, and lastly by confirming that the situation is under control.

### *Stimulation*

E1 stated that enjoyment is lacking because the types of questions a QAS attempts to answer are dissatisfiers by default; "if the answers are there, it will not necessarily be positive, but the absence of answers will definitely lead to dissatisfaction". However, it can be helpful to emphasize that a system is being used a lot and received positive feedback because it stimulates others to use it as well. The tone-of-voice can also be changed to add a bit of humor or bring things more lightly to unleash a smile, argues E4. Visualizations that accompany answers may be stimulating for users, if they are of value in relation to the answer. E5 argues that stimulation is also created by doing more than just answering the initial question; "try to follow-up the question, provide interpersonal contact options, alternative answers, or related articles from the organization".

## 7.2.3 Measuring user opinions

The experts did not mention the importance of measuring the users' opinions, but as stated, the experts might not have considered it part of realizing UX enrichment for QASs.

## 7.2.4 Other remarks

The experts also made statements on factors of influence on the experience that can be addressed to enrich a QAS which are not directly relatable to the topics found in the UX enrichment guidelines for QASs.

### *Satisfiers*

E3 argued that the user will feel good especially when the pragmatic quality enables the user to get an answer, because the user will probably anticipate that it will not work properly. "Especially when you're talking about this service context, the most important thing for the user is that the answer suits what he searched for, and the success of this can be seen as a hedonic quality. I assume that this system is not meant to be enriched". E5 suggested to look at emotional aspects to be intertwined in the system, for example, in relation to provided answers; "if the question is about a lost bankcard, you could respond with an emphatic message in addition to the answer, to spark a feeling of trust which can contribute to the relation with the bank". In line with that, E4 addressed that the system will be used by different types of people, something a CSR could notice and adjust to from an emotional perspective; "maybe the system can take that into account". E4 further claimed that answers should be presented clearly and as short as possible, possibly by using bullet points. People also get things from a page layout; "a calm and clear layout feels safe for people, while a page with a lot of images can spark a feeling of distrust, making the user question whether the page is the real deal".

### *Dissatisfiers*

E1 expressed that with the design of a system, never forget to make things very clear for a user; "for the designer, everything is as clear as can be, and subsequently, he often fails to make things clear to the user". In line with that is a remark of E4, stating that users not always know why things are how they are, and that creates a variation in perceptions. Some might have a positive perception, while others have a negative perception about the exact same matter due to the lack of clarity. About the response of the system to a question, E3 argued that for a QAS, it is probably bad for a person's feelings towards the system to respond with possible answers, because the system is apparently not able to properly understand your question. E4 made a remark on the wordings of the response when the questions was not answerable; "if you state that "we do not have an answer", it might give the user the feeling that we are a bunch of losers who cannot answer a question because "we" is associated with the organization as a whole, while someone of the organization can absolutely answer the question".

### *Pragmatic improvements*

As to be expected, some experts expressed possible improvements that are more related to the functionality or usability of the system (its pragmatic attributes/quality) rather than the hedonic side of things. E2 suggested that the question-bar should show suggestions while typing and to look at which answers are most helpful for which answers, instead of just returning the number one search results (answer candidate). Also, if no answer is found, traditional search results should appear. What also could help is to add tagging to answers based on the existing taxonomy, to show in which domain the answer was found. E5 suggested that it can be helpful to use knowledge of the user (i.e. customer), to find a relevant answer and to let the user know that we recognize him to gain trust. E5 addressed that different types of users should be considered for the possible forms of answers. "A user might understand that it is a system that cannot produce a natural dialogue, but another user might be convinced that he can have a natural dialogue and therefore expects answers in a natural manner". E5 furthermore questioned whether it is smart to aim for answering with a single answer, because the system might just fail to answer the question. It could be good to show other answer candidates as well. It might show the user that the system tries to think along with the user. E1 suggested that the system should end an interaction with feedback on performed actions and agreements, and what the user can expect next.

## 7.2.5  Concluding opinions

In the first phase, only E1 and E3 concluded in a more concluding sense about the UX enrichment. E1 expressed that some human needs are probably more applicable than others. E3 added to that by stating "if you start doing the things mentioned for enrichment in theory, you will create a lot which distracts from what this system is really about".

## 7.3 Expert opinions on the user experience enrichment guidelines for question-answering systems

In the second phase, the experts gave their opinions on the UX enrichment guidelines for QASs based on an understanding of the guidelines and its application through the enriched XBank QAS PoC. The results from this second phase are presented in this section. The paragraphs are again based on the node tree used for coding.

### 7.3.1 Managing user anticipations

This topic was praised a lot but had its flaws as well. The experts agreed that it is good to make clear what a user can expect from the system by explaining what it is capable of and also what can be submitted in its question-bar. "Hints are handy" (E2). However, it should be brief; "the feed-forward is good but keep it as short as possible otherwise it misses its purpose" (E5). E3 even stated that the example in the PoC is dangerous because it counts four sentences; "you get a wall-of-text, and the effect on the user will be like "screw that" and it won't be read". A suggestion is that such an extensive explanation may initially not be required but could be presented when the system failed to deliver an answer. For the question-bar, it should be investigated whether an indication of possible input is a good idea compared to just stating that a question can be asked in a natural way (E5). In addition, E5 emphasized that the question-bar is presented like a traditional search-bar, and that might not give the impression that the user is confronted with a QAS; "the user sees the magnifying glass and will assume it to be a search-bar".

### 7.3.2 Addressing human needs

This paragraph discusses the experts' comments on the guidelines and the enriched QAS PoC in relation to the human needs. Most comments were explicitly linked to these human needs by the experts but again, there were also comments that rather unintentionally touched upon these topics.

*Autonomy*
In general, the experts agreed that it is a good idea to present all options in a single screen. Especially compared to current practice, it is user friendly to present options (to get an answer) at a glance and enables to user to decide how to fulfill the process. E4 expressed that it is a good idea to present the question-bar above all texts so that the user can always start typing or either choose to read the explanation first. E5 agreed that it is a good thing to never impose a user to get in contact; "it should just be an option". E2 is however a bit skeptical with the visibility of contact options by referring to the so-called F-model; "are users not going to make massive use of the presented interpersonal contact options if you show it directly at the top, and thereby skip the QAS?". He predicts that users will move on to calling the organization rather than trying out the QAS. E3 adds to that by explaining that the 'hidden' contact options is a pure business decision; "if you show how they can get in contact while keeping the same number of CSRs, the queues will grow which is also a bad UX. That is why we placed contact options at the bottom of a service page, because we were not able to handle more incoming calls".

*Competence*
In general, the experts agreed with the competence guidelines. Repeating the user's question is nice because it states which exact question is being answered, however; "if you repeat the literal question, the answer should really suit it, otherwise this will raise eyebrows" (E4). Also, the complementing message when an answer is found can be perceived weirdly. In the PoC, this message states "score!" to celebrate that an answer has been found, but it is a bad response if the question is concerned with, for example, a stolen bankcard. E1 agreed by stating the same. E4 did praise the response in the case of an unanswerable question; "good that it states the system failed if the question is unanswerable, good to clarify that it is the system's fault".

*Popularity*
Most comments regarding popularity were related to the friendly-worded welcoming messages. While it is decent to welcome a user with "good afternoon", E2 advised to clarify whether it comes from a system, a chatbot, or an employee. E5 called this an example of a quick win; "it is an easy adjustment and doesn't require much space". E3 suggested that this personal 'page title' should replace the default page title, but also stated that in the case of a QAS on an organization's service page, it can be questioned whether you should only add such a

natural, welcoming message on the homepage. E5 again suggested that certain answers can be accompanied with an emphatic message because certain answers implicate that the user's situation is an unpleasant situation. This is also a quick win. The opinions on addressing the user as an expert when they enter a feedback screen differ; "fun idea, and good because it gives the user the feeling of being an expert" (E3), while E4 is of the opinion that it might be a bit too much and would like to research this; "if I would read "hi expert!" I would think "come on, I am just asking a question"". E4 also expressed that it is indeed important to give the user the impression that feedback is taken seriously but would like to have some sort of feedback towards the user as well; "the user will quickly think that it is useless to submit feedback, so think of a way to gain the user's trust that the feedback will truly be considered". Only if the user has the impression that he or others will benefit from the feedback, he will be open to provide some.

### *Relatedness*
Like in the first phase, relatedness is the least discussed topic. E4 really liked the idea of sharing the question and answer directly from the answer-screen. E5 expressed that the interpersonal contact options and its emphasis on the possibility of getting in contact with a CSR is good because it shows that the organization wants to keep the conversation going. It always offers the user a way out. Its placement on the screens should however always be the same for predictability reasons.

### *Security*
The message to spark a feeling of security was received well. E2 however stated that it can be a dangerous promise to state that every problem can be solved; "even if the organization is of the opinion that a problem is solved, the user might think different because a process might still be in progress while the organization's actions are done". The wording of the message can be adjusted but it is still a smart idea to comfort the user this way. E3 emphasized that the message of this security statement should already come forward from the title, for direct clarity and comfort. E4 noticed that the security statement shown in the response when a question is unanswerable is the exact same as the initial security statement; "it says "solving problems starts here" but we already started solving the problem before? Apparently, the system couldn't solve the problem just yet so I would advise to change the second statement and add the possibility of getting in contact with a CSR".

### *Stimulation*
For stimulation, the experts foremost discussed the idea of showing fun facts, and specifically the example of the number of helpful answers given (as exemplified in the PoC). Before, E1 already expressed that it is a good idea to show the use and the users' opinion of the system to stimulate others to use the system as well. E2 agreed that it is smart to express what the system is doing (i.e. providing helpful answers). However, the real numbers might be disappointing. E4 also expressed that right now, it seems like a nice sales pitch, because it does not show how many answers were not helpful; "maybe show a percentage, to show the chances of finding a helpful answer". However, E5 stated that both showing the amount and a percentage pose a risk because it might be disappointing and thus counterproductive. A suggestion was to show it afterwards when an answer is found, to stimulate future use. If it is shown before and the system fails to help the user, it will create a feeling of failure.

## 7.3.3  Measuring user opinions

E2 thinks it is a cool idea to let users provide feedback regarding their experiences and approaching users who have tried the system multiple times as an expert is good. E3 thinks this requires a little bit more attention; "it should not be possible for a user to be approached for feedback before having their question answered, because otherwise things will start to mix". E5 emphasized that the questionnaire should eventually be combined with the user's questions and answers provided, to get insights on the experience in relation to those elements rather than for the QAS in general. Regarding the content of the actual questionnaire that was formed, E4 had some clear critique and suggestions; "usually, people will be open to answering 10 questions, but 33 is too much. Decide which questions are most valuable and take out the ones that matter the least. For an organization, it should result in pragmatic insights rather than being scientific". Also, the wordings of questions should be extremely easy for users to understand; "if the question is hard to understand, users will leave". Especially the part based on PANAS should be shortened according to E4. Some questions are interesting, but others are nonsense; "if people drop out of the questionnaire, you have nothing, so it is better to get a few answers by keeping it short and use simple language". The part regarding pragmatic and hedonic quality was considered very relevant. The questions regarding the user type might result in fake results because users might not admit what

they really are. Questions regarding the context are good because the stated context factors do influence the experience. E5 also commented that the current questionnaire is very long and should definitely be shortened. E3 stated that, although it is important to measure which types of people and which situations are linked to which human needs, not everything can be truly measured; "it is inevitable to do some assumptions based on your own insights".

## 7.3.4 Other remarks

The experts still made a few statements on factors of influence on the experience which are not directly relatable to the topics found in the UX enrichment guidelines for QASs. However, no additional satisfiers were mentioned.

### *Dissatisfiers*
Two points of attention were brought forward by E4. It was suggested that the question-bar on an answer-screen should not state to "try" another question, because this implicates that it is shooting with buckshot, hoping to get lucky in finding some answer. Rather just use the word "ask". A few more comments were made on the wordings used in texts, suggesting being careful with it because words influence interpretations. Furthermore, it was stated that approaching users for their opinion on the system's experience after multiple usages might spark a feeling of intrusiveness, which is unpleasant; "a user might think that you are watching them closely" (E4).

### *Pragmatic improvements*
E2 stated that recognition is an important aspect; "create an appearance similar to what people are used to in other applications". Furthermore, E1 suggests showing suggestions while typing in the question-bar to help a user formulate his question. E1 also advised to act when a user clicks "no" (in response to whether the answer was helpful or not). Instead of asking for feedback, the user should rather be assisted in finding a helpful answer because the "no" is a clear sign that the user's problem still exists. E5 suggested that repeating the user's question can be improved by reformulating the question as if it is reacted to, for example, by stating "you asked a question about …". Again, E5 suggests accompanying answers with alternative answers (i.e. other answer candidates) because the answer might not yet fulfill the user's information need. Both E2 and E5 also expressed that visual improvements are possible, especially to offer all options appropriately for improved effectivity.

## 7.3.5 Concluding opinions

In the second phase, the experts were able to reflect more generally on the UX enrichment guidelines for QASs and the enriched XBank QAS PoC, and thus a lot of more overarching, concluding comments were given. E2 stated that the guidelines can definitely lead to a better experience for users and thinks that the proposed areas of attention are nice and interesting; "especially because this QAS is a functionality on a service page, a user will not have a clear initial opinion about the system, and that is why the possible UX enrichments are important". E3 stated that all the enrichments are interesting and pleasant to see. E1 stated that the guidelines should be used carefully because it could lead to a "Christmas tree" of options, and that could cause options to be skipped by the user. "Good work, my compliments. However, less is more. The system should really be about the question-bar, and people are used to using Google. Useful options surrounding the question-bar is user-dependent and you should find out what a user really misses to make choices for enrichment" (E1). E3 also addressed that the enrichment is a bit too much which makes it difficult to maintain focus; "what you did, working down a theoretical list, is often what organizations do, but you see what happens", referring to the excessive enriched XBank QAS PoC. However, E3 stated that doing everything is emphasized as a good starting point because it enables to make choices to get to an optimum. Only the effective elements should be integrated. E5 states that the human needs are all understandable but also emphasizes that the overall result of their application leads to the conclusion that the system's core goal might be overseen. It became a bit too much and especially the guiding texts are too extensive; "I think you have to make choices to make the interaction efficient, by prioritizing the elements and define them as primary and secondary, or even superfluous". E4 agrees by stating that the guidelines could lead to a very extensive enrichment and it is questionable whether users really appreciate this; "this is something you can test with actual users by observing their behavior while letting them try out the system, and this will result in other insights compared to this analysis through expert reviews". E3 explained this by stating that "each element should get its own cost-benefit analysis". E5 also suggested to research which elements are most effective and would recommend not to apply the full palette; "you can share an answer, indicate whether the answer was helpful, provide feedback; all those options might overshadow the

goal of the system. An option is to pick one human need and try out variations to find out what is important, and then define the priority of options to really make a difference". E3 also suggested to do more research with more examples and variations in which clear choices are made, because doing everything is clearly too much; "choosing one or two things will reduce the amount of texts, increasing the chances of being read, and it would be best to intertwine it in texts that are already there". A lot of things can be combined with each other, integrated in existing elements, or simply shortened, to prevent that the user must think about everything shown on screen. E3 finally concluded with a summarizing statement by saying that "it is good to make the distinction between the pragmatic and hedonic part, but the hedonic part shouldn't get in the way of the pragmatic part of the system".

# 8 DISCUSSION

This chapter first discusses the interpretations of the results, followed by a reflection on the unfolded research process, limitations of the research, and lastly, suggested directions for future research are described.

## 8.1 Interpretations of the results

In this section, the analysis on the discussed results is presented by reflecting on the expert opinions. All findings are recapitulated to conclude about the two objectives of the validation: (1) the similarities and differences between the expert opinions on enrichment and the designed UX enrichment guidelines for QASs, and (2) the adequacy and shortcomings of the designed UX enrichment guidelines for QASs. The combination enables to reflect on the main research objective; how a QAS can be enriched with UX concepts.

### 8.1.1 Reflecting on the expert opinions on how to enrich a common practice question-answering system

The results from the first phase showed that the ideas of experts on enriching a QAS are much in line with the designed UX enrichment guidelines for QASs, as many of the guidelines' topics were touched upon. In this paragraph, both the corresponding and deviating ideas and opinions are briefly discussed per topic to recapitulate and then conclude what the first validation phase has proofed.

*Experts' suggestions for user experience enrichment in comparison to the guidelines*
  ➢ The experts agreed that managing the users' anticipations is an important area of attention, and the contents of both the guidelines were addressed. Additional suggestions were to think about the system's appearance as well, to visually manage what a user expects from the functionality, and it could be more clearly be expressed that the system is autonomous (i.e. not an interpersonal interaction).
  ➢ Comments that addressed possible enrichments were analyzed to clarify to what extent the guidelines for addressing human needs are similar to expert opinions. Only for autonomy, a complete match between expert opinions and the guidelines is found. For competence, the expert opinions were mostly in line. They only did not mention to repeat the user's question. An additional suggestion was to not only take the blame when the system failed to deliver an answer, but also explain why this happened. For popularity, only the possibility for providing feedback was suggested, with the condition that it is not forced upon the user to provide feedback. Another suggestion was to express appreciation for the user for using the QAS. Relatedness was not quite in line. Relatedness between the user and close ones was not mentioned at all. Only the importance of relatedness between the user and the organization was addressed. An emphasis on possible interpersonal contact was mentioned, but only in the case of an unanswerable question. Security was in line in terms of comforting a user only, but it was found specifically important when panicky situations occur. It was suggested that for unanswerable questions, accusations towards the user should be avoided to prevent insecurity. Security was argued to be absent in the QAS PoC, and one expert argued that it would not be that relevant. For stimulation, the expert opinions were in line with the guidelines. In addition, it was suggested to see if visualizations can be used as well. One expert claimed that the system is not meant for enjoyment because of its service context, which is by default a dissatisfying experience because of the existence of a question or problem.
  ➢ No comments were given regarding the measurement of user opinions of the system's experience.
  ➢ Some of the experts' ideas about satisfiers and dissatisfiers were not directly relatable to the discussed topics. Areas touched upon were enrichment with emotional aspects (possibly in relation to the asked question or the type of user), the form of answers, the page layout, (lack of) clarity, (lack of) a fitting response, and (wrong) use of particular words. On the pragmatic side of the system, several interesting improvements were suggested (which might indeed enhance the experience) but are not considered an enrichment from a UX perspective.
  ➢ In a more concluding sense, it was indicated that certain enrichments are more applicable than others, and that the goal should not be to cover everything because it will lead to distractions.

### *Conclusion: similarities and differences*

Analyzing these results leads to the conclusion that with knowledge and understanding of the research background and a common practice QAS PoC, the experts can come up with partially corresponding guidelines to enrich a QAS from a UX perspective. It must be stated that only the opinions of the five experts combined realized a nice palette of suggestions for guidelines (i.e. individually, none was able to do so). At certain points, the experts brought more extensive ideas forward (compared to the formulated guidelines) and a few additional ideas were suggested as well (i.e. justified for a certain human need but absent in the guidelines). Purposes of some guidelines were however only partially addressed or not found at all in the experts' comments. The idea of measuring user opinions was not mentioned, but this can be clarified by assuming that measurement is not part of the actual enrichment; it is only a method to get insights about the enrichments. Lastly, additional ideas were suggested that are considered out of the defined UX scope or classified as pragmatic improvements. While these ideas might indeed enhance the experience, they are not explicitly in scope for UX enrichment guidelines for QASs. In conclusion, it was advised that enrichments should be limited due to their potential distractive effect.

## 8.1.2 Reflecting on the expert opinions on the user experience enrichment guidelines for question-answering systems

The results from the second phase showed that although the designed UX enrichment guidelines for QASs are well received, they have clear shortcomings. In this paragraph, the guidelines agreed upon, suggestions for improvements, and the shortcomings of the guidelines are briefly discussed per topic to recapitulate and then conclude what the second validation phase has proofed.

### *Experts' opinions on the user experience enrichment guidelines for question-answering systems*

➢ The experts agreed with the guidelines for managing user anticipations and praised this part. However, explaining texts should really be as short as possible, while an extensive explanation could be effective in the response to an unanswerable question. The optimal form of the explaining placeholder text in the question-bar should be investigated, and the question-bar's appearance could be changed for a distinction in comparison to a traditional search-bar.

➢ Comments on the guidelines for addressing human needs and their accompanying examples in the QAS PoC were analyzed to define its strengths and flaws. For autonomy, the guidelines were praised, foremost for its emphasis on giving the user control and preventing imposition. While it is beneficial for a feeling of autonomy to visually show all possible (interpersonal contact) options, it might cause users to skip the QAS and create an overload of users trying to get in contact with CSRs. For competence, the guidelines were complimented with only a few remarks. When a user question is repeated, the answer should also fit that question, and the celebrating message (after finding an answer) should be adapted to the answer context (i.e. cheerful or comforting). The friendly-worded welcoming message for a feeling of popularity was praised and characterized as a quick win, but it should be clear that it is a system who 'says' this, and it should be the main page title rather than an additional message. Empathic messages in unpleasant situations (i.e. derivable from the answer) might also help. Approaching users for feedback is also praised but the effect of approaching them as an 'expert' is not clear since experts' opinions on it differ. Also, asking users for feedback should be accompanied with the assurance that something (beneficial) will happen with it, although it is not clear how to achieve that. For relatedness, both guidelines were confirmed to be useful. A little remark was that the offered interpersonal contact options should however have a static placement on the various screens for predictability reasons. The security guidelines were also complimented, but small changes in the wordings can make a difference and it is preferred to propagate a message of security more directly rather than with a separate message. Also, the security messages in different screens should not be identical. The guideline for a feeling of stimulation through fun facts was heavily discussed. It can really stimulate users to use the system as well, but it might also uncover the system's low performance (if it is often not helpful) or give users a feeling of incompetence if they submit an unanswerable question.

➢ While it is praised to measure user opinions regarding the UX, the proposed questionnaire needs to be revised; the number of questions needs to be reduced drastically by selecting only the most relevant and insightful questions, and the formulation of questions should be as simple as possible. Both these aspects will make diminish people's aversion towards completing the questionnaire. The different components of the questionnaire were all confirmed to be useful. It was suggested to combine questionnaire results with the user's actions performed with the system. And while measuring is good,

it is impossible to precisely measure every aspect; own insights should sometimes be considered as well despite being based on assumptions.

➢ A few remarks were made outside of the scope of the guidelines. Certain wordings might be dissatisfying, and a feeling of intrusiveness might be sparked with an approach for feedback (because it confirms that the system knows of the interactions). Furthermore, a few pragmatic improvements (functionally and visually) were suggested based on what the XBank QAS PoC conveyed.

➢ The experts were able to reflect on the complete guidelines and the enriched XBank QAS PoC which exemplified what the guidelines could lead to. A very strong shared opinion was the result: following all the guidelines leads to an excessive enrichment, overshadowing the core goal of the system. The guidelines can however doubtlessly be used to realize effective enrichments to enable a good experience for users, and it is therefore key to find a balance with the possible (combinations of) enrichments. It is a good starting point to think of (or prototype) all options but choices should be made eventually, preferably after more research to prioritize possible (combinations of) enrichments based on their effectiveness.

### *Conclusion: adequacy and shortcomings*

Analyzing these results leads to the conclusion that with full knowledge and understanding of the designed guidelines and the exemplifying enriched QAS PoC, the experts agree with a large part of the guidelines. Also, the experts' opinions do not differ much from each other. For all the guidelines regarding managing user anticipations and addressing human needs, the experts did suggest adjustments, but the overall opinion was that these guidelines prescribe enrichments that make sense and could be effective. Measuring user opinions was praised, but the proposed questionnaire was firmly criticized and must be revised because it is not user-ready due to its length and wordings. Furthermore, only a few little suggestions were provided that are considered out of the defined UX scope of classified as pragmatic improvements. In conclusion, the experts heavily shared the opinion that the guidelines will lead to an excessive enrichment if they are as literally realized as exemplified in the enriched QAS PoC. Therefore, it is advised to determine a prioritization of enrichments, preferable through more research on the possible enrichments: their form, possible combinations, and foremost their effectiveness.

## 8.1.3 Final conclusions

To reflect on the main research objective, final conclusions are drawn regarding the expert opinions. As expected, due to their professional background, the experts had similar and sometimes even more extensive ideas for the enrichment of a QAS in terms of managing anticipations and addressing human needs, as well as some additional ones. Their 'out of scope'-ideas were to be expected as well, since their thought process was not explicitly bound by the research scope. However, those are not considered explicit UX enrichments and thus not considered as missing in the designed guidelines. The fact that none of the experts suggested measuring user opinions is not surprising because it is not a literal enrichment in the QAS that enhances the user's experience directly. The first phase finally led to suggestions emphasizing that enrichments should be limited to prevent a distractive effect, and this already clarified the most important misconception in the treatment design: enrichment of a QAS seemed somewhat endless. Therefore, the guidelines provide enrichment instructions on multiple, specific areas but do not instruct on the amount of enrichments to be applied. The results of the second phase, in which the experts reacted directly on the guidelines and the QAS PoC that exemplified its application, were therefore for a large part directed at this shortcoming. All experts agreed about the QAS being too excessively enriched due to applying all the guidelines and a lack of instructions on balancing the enrichments. This is assumed to be counterproductive for the system's usage and its experience. However, the individual guidelines were all agreed upon by the experts; their usefulness is confirmed, and their effectiveness can be assumed to a certain extent. As to be expected, improvements were suggested nevertheless, indicating little shortcoming for several guidelines. No additional UX-specific guidelines were suggested indicating that the set of guidelines might represent a complete set of UX enrichments for a QAS. Because of the chances of too excessive enrichment, research must be conducted to experiment with variations and combinations of enrichments to determine the most effective QAS enrichments. A prioritization of enrichments must be determined. Regarding the UX questionnaire that was created, experts were critical because of its length and wordings of the questions. This was to be expected because the questionnaire was mainly based on theoretical constructs used in extensive research on experiences of interactions with technology. Actual users are different than subjects that consciously participate in such research, and therefore, the questionnaire regarding the UX should be adjusted for higher chance of acceptance by users in practice.

## 8.2    Reflecting on the research process

A brief reflection on the overall research process explains how the research unfolded and led to the present results. The research was originally intended to result in a design of a QAS and a functional prototype following that design to demonstrate the possibility of automated handling of frequently asked customer questions with UX enrichment. Therefore, the technical aspects of a QAS were extensively explored during the literature review, including the possibilities of processing Dutch natural language and ways of validating the QAS performance. The literature review on UX however steered at non-functional aspects of systems rather than their functional: the distinction between pragmatic and hedonic. This was also touched upon in the scarce literature on the experience-side of QASs. While the pragmatic quality of systems is of importance, the hedonic quality is proofed to be more strongly tied with the experience a user has with a system, especially a positive experience. It was therefore chosen to focus on enrichments in the area of non-functional system aspects. This also led to the execution of an explorative observation of existing online QASs of Dutch commercial organizations as part of the empirical investigation. This resulted in an outline of the structure of a common practice Dutch online QAS. The literature review on UX was also extended with empirical results through unstructured interviews with people with knowledge of customer service and its IT solutions. This foremost confirmed the findings from the literature review and led to emphasis on certain aspects rather than contradicting them. It was chosen to continue with the scientific theories on UX as a foundation for the design of the treatment which is purposed to overcome the lack of knowledge regarding the UX in QAS designs: a set of guidelines that instruct how a QAS can be enriched from a UX perspective. The common practice Dutch online QAS structure was used as the reference model for a QAS that needs to be enriched. This, because of the initial research scope of researching QASs in a customer service context of a Dutch commercial organization. The main guideline topics for the enrichment of QASs were based on the insights from the UX literature and empirical insights: managing user anticipations, addressing human needs, and measuring user opinions. Since explicit UX guidelines are absent in literature, guidelines were developed from scratch: for each guideline, an explanation was written, accompanied by an explicit goal and an example of an enrichment in a QAS if the guideline would be applied. This resulted in a full set of guidelines which was named the *UX enrichment guidelines for QASs*. In addition, generic QAS PoCs were developed to visually exemplify what changes the guidelines could achieve. To validate the guidelines, it was chosen to perform a validation by expert opinions. Relevant experts within XBank were selected, and individual validation sessions were organized. To help the experts imagining the application of the guidelines, it was decided to develop additional QAS PoCs that had an XBank context. This, to collect more realistic predictions about the effects of the guidelines compared to predictions based on the generic QAS PoCs. A validation procedure was defined to ensure that each session resulted in analyzable qualitative data. Two sets of expert opinions were gathered. The first set consists of the experts' unbiased opinions on how to enrich a QAS from their perspective as a professional, to conclude whether similar or different ideas for enrichments would be suggested. The second set consists of the experts' opinions on the actual guidelines, to conclude on its adequacy and shortcomings. The combination of these conclusions enabled to objectively reflect on the designed treatment.

## 8.3   Limitations

Three clear limitations can be addressed that limited the research and its conclusions. Those three limitations and their consequences are briefly described including an explanation for their existence.

➢ Although a realistic, interactive mock-up was created to demonstrate the QAS as if it was functional, it was nowhere near a functional prototype of a QAS. This did limit the research possibilities. It limited the variation of use cases because each screen was a static image: it was not possible to type a question and get an actual response. Therefore, only the three scenarios of an answerable, unspecific, and unanswerable user question were implemented (with example questions and responses). The consequence of this is that it was not possible to do performance measurements to take the system's pragmatic quality into account. More importantly, it was not possible to conduct experiments with (fictional) users. Therefore, real experiences of users were absent in this research. This made it impossible to conclude about the actual effects of enrichments, and the relations with user types and user contexts remained uncovered as well. Instead, it was only possible to demonstrate the interactive mock-ups which is why validation by expert opinions was the used validation method. This was however a deliberate choice because it was beyond the scope of this research to develop a functional QAS prototype. This, due to conclusions drawn from the UX literature (i.e. pragmatic vs. hedonic) but also because of the amount of knowledge, work, and time it would require for its development and subsequently conduct experiments with enough (fictional) users for enough quantitative data.

➢ Another limitation is the context for which this research was conducted: a customer service context of a Dutch commercial organization (with XBank in particular) for the QAS PoCs that were used for the validation procedure. In this research, it was not possible to encompass the entire range of QAS application domains, and therefore, the generalization of the results can be weak but. However, it was a deliberate choice because it did create a very clear scope for the research. It also led to a more realistic and familiar context for the experts, bettering their imaginations of the treatment in the problem context and thus more reliable expert opinions.

➢ Lastly, it can also be seen as a limitation that the experts only imagined the application of the UX enrichment guidelines for QASs. The experts did not use the guidelines themselves to experience how helpful they are in the process of designing or enriching a QAS. While this would result in different insights and opinions on the guidelines, the validation by expert opinions is however intended to result in qualitative data by using the imagination of experts. Therefore, it was not a wrong validation, but the reader should bear in mind that the conclusions of this research are based on the experts' imaginations.

## 8.4    Future research

Future research is necessary to extend the knowledge on UX enrichment in QASs. This research was a first attempt at outlining how to enrich a QAS with UX concepts. Three possible future research directions are described.

> ➢ As concluded in this research, it is required to determine the effectiveness of (combinations of) the particular enrichments to prioritize enrichments. Future research should be carried out using a functional QAS prototype or an existing QAS that can be enriched. This, to apply the UX enrichment guidelines for QASs in a functional QAS and let users experience variations of an enriched QAS. This can be in an experimental setup (i.e. with fictional users) or with live users. The goal of such research is to measure what users experience to determine the true effects of various enrichments. The importance of pragmatic quality can then also be taken into account. It also provides an opportunity to define an effective UX questionnaire in a form that is accepted by users (i.e. which they are willing to fill in) that measures the concepts currently conveyed in the UX questionnaire. Besides the effects of various enrichments, it is also of interest to link the effects to different types of users and types of contexts. The ultimate goal is to design QASs with enrichments that are moderated to the user and its context.

> ➢ Future research related to UX should be undertaken to determine whether the six human needs form the definite set of human needs to be conveyed in a QAS for realization of hedonic quality. Originally, ten basic human needs were defined of which six were selected to be most important in interactions with technology. This was already argued not to be a definitive list and therefore it is of interest to determine whether, for example, the human need 'meaning' could also be a motive for certain enrichments. Also, more out-of-the-box concepts could be researched that may be relevant in the context of a QAS, moderated to the user and its context. For example, 'intrusiveness' may be an interesting concept to consider, especially with the moderations based on user type and/or context.

> ➢ Another possible direction for future research is the exploration of technical possibilities for user type and context determination and possible moderations based on that. If it is possible to determine a certain user type and/or context based on, for example, tracking or cookies of a website visitor (before a possible login) or its profile and performed actions based on existing knowledge of the person (after login), the QAS could convey personalized and more relevant enrichments as well. Future research could explore these concepts.

# 9 CONCLUSION

In research on QASs, attention for the experience of using such systems is lacking. Meanwhile, XBank is aspiring to automate question-answering but it is unclear how to ensure the '9+ customer experience' that they strive for with their question-answering through interpersonal contact. Therefore, research was conducted at XBank to answer the following research question:

*How can a question-answering system be enriched with user experience concepts
to better handle frequently asked customer questions?*

A literature review and an empirical investigation were conducted to explore the body of knowledge on QASs and UX. This resulted in an understanding of QASs in terms of technical configurational possibilities to automate question-answering and structures of current QAS implementations at various Dutch commercial organizations for customer service purposes. Furthermore, it resulted in an understanding of UX in terms of its true meaning, the elements that play a role in unfolding experiences, and how positive UX can be enabled. Based on these insights, the treatment that had to be designed to enrich QASs with UX concepts was outlined: *UX enrichment guidelines for QASs*. These guidelines had to cover means to support the user's be-goals, ensure that the user has realistic anticipations, and provide the organization with insights. Therefore, guidelines were defined to manage user anticipations, address human needs, and measure user opinions. This explains how UX concepts can be covered in a QAS. In addition, generic QAS PoCs were realized to exemplify the application of these guidelines. A validation by expert opinions was conducted to validate the guidelines, uncovering its adequacy and shortcomings. Opinions on the guidelines were gathered from five XBank employees, who were classified as experts based on their knowledge of UX and familiarity with XBank's customer service domain. XBank QAS PoCs were realized in addition for the validation procedure to exemplify the guidelines' application in a realistic and familiar context. Reflecting on the findings of the validation leads to the final conclusion that, according to the experts, the guidelines can lead to useful, effective enrichments for a QAS in terms of UX because relevant aspects are included, and the guidelines clearly explain how these enrichments can be conveyed. However, the guidelines are yet to be optimized and are currently not ready for use in practice. The main issue indicated by the experts is the lack of instructions on balancing the enrichments to prevent a too excessively enriched QAS, of which it is predicted to cause a counterproductive effect on the system's usage and its experience. More research on the effectiveness of particular (combinations of) enrichments must be conducted to determine a prioritization of enrichments and eventually an enrichment optimum.

# BIBLIOGRAPHY

Abdi, A., Idris, N., & Ahmad, Z. (2018). QAPD: an ontology-based question answering system in the physics domain. *Soft Computing, 22*(1), 213-230.

Allam, A. M., & Haggag, M. H. (2012). The question answering systems: A survey. *International Journal of Research and Reviews in Information Sciences, 2*(3).

Athira, P. M., Sreeja, M., & Reghuraj, P. C. (2013). Architecture of an ontology-based domain-specific natural language question answering system. *International Journal of Web & Semantic Technology, 4*(4), 31-39.

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit.* Sebastopol: O'Reilly Media, Inc.

Booij, G. (2012). *The grammar of words: An introduction to linguistic morphology.* Oxford: Oxford University Press.

Brill, E., Dumais, S., & Banko, M. (2002). An analysis of the AskMSR question-answering system. *Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10* (pp. 257-264). Association for Computational Linguistics.

Brunnström, K., Beker, S. A., De Moor, K., Dooms, A., Egger, S., Garcia, M. N., . . . Larabi, M. (2014). Qualinet white paper on definitions of quality of experience. Novi Sad. Retrieved from https://hal.archives-ouvertes.fr/hal-00977812

Campr, M., & Ježek, K. (2015). Comparing semantic models for evaluating automatic document summarization. *International Conference on Text, Speech, and Dialogue* (pp. 252-260). Cham: Springer.

Cao, Y., Liu, F., Simpson, P., Antieau, L., Bennett, A., Cimino, J. J., . . . Yu, H. (2011). AskHERMES: An online question answering system for complex clinical questions. *Journal of biomedical informatics, 44*(2), 277-288.

Chaffey, D., Ellis-Chadwick, F., Mayer, R., & Johnston, K. (2009). *Internet marketing: strategy, implementation and practice.* Harlow: Pearson Education.

Chen, D., Fisch, A., Weston, J., & Bordes, A. (2017). Reading wikipedia to answer open-domain questions. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. 1*, pp. 1870-1879. Vancouver: Association for Computational Linguistics.

Diekema, A. R., Yilmazel, O., & Liddy, E. D. (2004). Evaluation of restricted domain question-answering systems. *School of Information Studies: Faculty Scholarship, 52*. Retrieved from School of Information Studies: Faculty Scholarship: https://surface.syr.edu/istpub/52

Ding, D. X., Hu, P. J., & Sheng, O. R. (2011). e-SELFQUAL: A scale for measuring online self-service quality. *Journal of Business Research, 64*(5), 508-515.

Dix, A., Finlay, J., Abowd, G. D., & Beale, R. (2004). *Human-Computer Interaction.* Harlow: Pearson Education Limited.

Dixon, M., Freeman, K., & Toman, N. (2010). Stop trying to delight your customers. *Harvard Business Review, 88*(7), 116-122.

Dumais, S., Banko, M., Brill, E., Lin, J., & Ng, A. (2002). Web question answering: is more always better? *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 291-298). ACM.

Dwivedi, S. K., & Singh, V. (2013). Research and reviews in question answering system. *Procedia Technology, 10*, 417-424.

Gentile, C., Spiller, N., & Noci, G. (2007). How to sustain the customer experience: An overview of experience components that co-create value with the customer. *European management journal, 25*(5), 395-410.

Gupta, P., & Gupta, V. (2012). A survey of text question answering techniques. *International Journal of Computer Applications, 53*(4), 1-8.

Hassenzahl, M. (2003). The Thing and I: Understanding the Relationship Between User and Product. In M. A. Blythe, K. Overbeeke, A. F. Monk, & P. C. Wright, *Funology: From Usability to Enjoyment* (pp. 31-42). Dordrecht: Kluwer Academic Publishers.

Hassenzahl, M. (2008). User Experience (UX): Towards an experiential perspective on product quality. *Proceedings of the 20th Conference on Human-Computer Interaction* (pp. 11-15). ACM.

Hassenzahl, M. (2010). Experience design: Technology for all the right reasons. In J. M. Carroll, & J. M. Carroll (Ed.), *Synthesis lectures on human-centered informatics* (Vol. 3, pp. 1-95). Essen: Morgan & Claypool.

Hassenzahl, M. (2013). User experience and experience design. In M. Soegaard, & R. F. Dam, *The Encyclopedia of Human-Computer Interaction.* Aarhus: Interaction Design Foundation.

Hassenzahl, M., & Monk, A. (2010). The inference of perceived usability from beauty. *Human–Computer Interaction, 25*(3), 235-260.

Hassenzahl, M., & Tractinsky, N. (2006). User experience - a research agenda. *25*(2), 91-97.

Hassenzahl, M., Diefenbach, S., & Göritz, A. (2010). Needs, affect, and interactive products – Facets of user experience. *Interacting with computers, 22*(5), 353-362.

Hassenzahl, M., Wiklund-Engblom, A., Bengs, A., Hägglund, S., & Diefenbach, S. (2015). Experience-oriented and product-oriented evaluation: psychological need fulfillment, positive affect, and product perception. *International journal of human-computer interaction, 31*(8), 530-544.

Higashinaka, R., & Isozaki, H. (2008). Automatically acquiring causal expression patterns from relation-annotated corpora to improve question answering for why-questions. *ACM Transactions on Asian Language Information Processing, 7*(2), 6.

Hovy, E. H., Gerber, L., Hermjakob, U., Junk, M., & Lin, C. Y. (2000). Question Answering in Webclopedia. *TREC, 52*, pp. 53-56.

ISO 9241-210:2010. (2010). *Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems.* Switzerland: International Organization for Standardization.

Jijkoun, V., Mishne, G., & de Rijke, M. (2003). Building infrastructure for Dutch question answering. *Proceedings of the Fourth Dutch-Belgian Information Retrieval Workshop*, (pp. 22-27).

Johnson, D. S., Bardhi, F., & Dunn, D. T. (2008). Understanding how technology paradoxes affect customer satisfaction with self-service technology: The role of performance ambiguity and trust in technology. *Psychology & Marketing, 25*(5), 416-443.

Kitchenham, B. (2004). *Procedures for performing systematic reviews.* Keele: Keele University.

Kolomiyets, O., & Moens, M. F. (2011). A survey on question answering technology from an information retrieval perspective. *Information Sciences, 181*(24), 5412-5434.

Korber, M., Eichinger, A., Bengler, K., & Olaverri-Monreal, C. (2013). User experience evaluation in an automotive context. *Intelligent Vehicles Symposium Workshops (IV Workshops)* (pp. 13-18). IEEE.

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (2007). *Handbook of latent semantic analysis.* New York: Psychology Press.

Law, E. L., Roto, V., Hassenzahl, M., Vermeeren, A. P., & Kort, J. (2009). Understanding, scoping and defining user experience: a survey approach. *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 719-728). Boston: ACM.

Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. *Proceedings of the 31st International Conference on International Conference on Machine Learning. 32*, pp. 1188-1196. Beijing: JMLR.

Lemke, F., Clark, M., & Wilson, H. (2011). Customer experience quality: an exploration in business and consumer contexts using repertory grid technique. *Journal of the Academy of Marketing Science, 39*(6), 846–869.

Lin, J. S., & Hsieh, P. L. (2011). Assessing the self-service technology encounters: development and validation of SSTQUAL scale. *Journal of Retailing, 87*(2), 194-206.

Lin, J., Quan, D., Sinha, V., Bakshi, K., Huynh, D., Katz, B., & Karger, D. R. (2003). What makes a good answer? The role of context in question answering. *Proceedings of the Ninth IFIP TC13 International Conference on Human-Computer Interaction* (pp. 25-32). Zurich: IOS Press.

Menger, V., Scheepers, F., & Spruit, M. (2018). Comparing Deep Learning and Classical Machine Learning Approaches for Predicting Inpatient Violence Incidents from Clinical Text. *Applied Sciences, 8*(6), 981.

Meuter, M. L., Ostrom, A. L., Roundtree, R. I., & Bitner, M. J. (2000). Self-service technologies: understanding customer satisfaction with technology-based service encounters. *Journal of marketing, 64*(3), 50-64.

Meyer, C., & Schwager, A. (2007). Customer experience. *Harvard business review, 85*(2), 116-126.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems 26. 1*, pp. 3111-3119. Lake Tahoe: Curran.

Minaee, S., & Liu, Z. (2017). Automatic question-answering using a deep similarity neural network. *IEEE Global Conference on Signal and Information Processing* (pp. 923-927). IEEE.

Mishra, A., & Jain, S. K. (2016). A survey on question answering systems with classification. *Journal of King Saud University-Computer and Information Sciences, 28*(3), 345-361.

Missen, M., Boughanem, M., & Cabanac, G. (2013). Opinion mining: reviewed from word to document level. *Social Network Analysis and Mining, 3*(1), 107–125.

Mollá, D., & Vicedo, J. L. (2007). Question answering in restricted domains: An overview. *Computational Linguistics, 33*(1), 41-61.

Ong, C. S., Day, M. Y., & Hsu, W. L. (2009). The measurement of user satisfaction with question answering systems. *Information & Management, 46*(7), 397-403.

QSR International. (2019, March 6). *What is NVivo? | NVivo*. Retrieved from NVivo qualitative data analysis software | QSR International: https://www.qsrinternational.com/nvivo/home

Ray, G., Muhanna, W. A., & Barney, J. B. (2005). Information technology and the performance of the customer service process: A resource-based analysis. *MIS quarterly, 29*(4), 625-652.

Roto, V., Law, E., Vermeeren, A. P., & Hoonhout, J. (2011). User Experience White Paper. Retrieved from http://www.allaboutux.org/files/UX-WhitePaper.pdf

Samarakoon, L., Kumarawadu, S., & Pulasinghe, K. (2011). Automated question answering for customer helpdesk applications. *6th International Conference on Industrial and Information Systems* (pp. 328-333). Sri Lanka: IEEE.

Sheldon, K. M., Elliot, A. J., Kim, Y., & Kasser, T. (2001). What is satisfying about satisfying events? Testing 10 candidate psychological needs. *Journal of personality and social psychology, 80*(2), 325-339.

Simmons, R. F. (1965). Answering English questions by computer: a survey. *Communications of the ACM, 8*(1), 53-70.

Sneiders, E. (2009). Automated FAQ answering with question-specific knowledge representation for web self-service. *2nd Conference on Human System Interactions* (pp. 298-305). Catania: IEEE.

Song, W., Feng, M., Gu, N., & Wenyin, L. (2007). Question similarity calculation for FAQ answering. *Third International Conference on Semantics, Knowledge and Grid* (pp. 298-301). Shan Xi: IEEE.

spaCy. (2018, October 9). *Dutch · spaCy Models Documentation*. Retrieved from spaCy · Industrial-strength Natural Language Processing in Python: https://spacy.io/models/nl

spaCy. (2018, October 9). *Language Processing Pipelines · spaCy Usage Documentation*. Retrieved from spaCy · Industrial-strength Natural Language Processing in Python: https://spacy.io/usage/processing-pipelines

Sportiche, D., Koopman, H., & Stabler, E. (2013). *An introduction to syntactic analysis and theory.* Chichester: John Wiley & Sons.

Tulkens, S., Emmery, C., & Daelemans, W. (2016). Evaluating unsupervised Dutch word embeddings as a linguistic resource. *Proceedings of the Tenth International Conference on Language Resources and Evaluation* (pp. 4130-4136). Paris: European Language Resources Association.

van der Sloot, K., van den Bosch, A., & van Gompel, M. (2018, October 9). *Frog*. Retrieved from Frog: https://languagemachines.github.io/frog/

Verhoef, P. C., Lemon, K. N., Parasuraman, A., Roggeveen, A., Tsiros, M., & Schlesinger, L. A. (2009). Customer experience creation: Determinants, dynamics and management strategies. *85*(1), 31-41.

Wang, M. (2006). A survey of answer extraction techniques in factoid question answering. *Computational Linguistics, 1*(1).

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology, 54*(6), 1063-1070.

Wieringa, R. (2014). *Design Science Methodology for Information Systems and Software Engineering.* London: Springer.

Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. *Proceedings of the 18th international conference on evaluation and assessment in software engineering.* ACM.

Yen, H. R. (2005). An attribute-based model of quality satisfaction for internet self-service technology. *The Service Industries Journal, 25*(5), 641-659.

Yogish, D., Manjunath, T. N., & Hegadi, R. S. (2016). A Survey of Intelligent Question Answering System Using NLP and Information Retrieval Techniques. *International Journal of Advanced Research in Computer and Communication Engineering, 5*(5), 536-540.

Zhou, J., & Xu, W. (2015). End-to-end learning of semantic role labeling using recurrent neural networks. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. 1*, pp. 1127-1137. Beijing: Association for Computational Linguistics.

Zhu, Y. Q., & Chen, H. G. (2015). Social media and human need satisfaction: Implications for social media marketing. *Business horizons, 58*(3), 335-345.

Zimmerling, E., Höllig, C. E., Sandner, P. G., & Welpe, I. M. (2019). Exploring the influence of common game elements on ideation output and motivation. *Journal of Business Research, 94*, 302-312.

# APPENDICES

## Appendix A: Concepts related to the experience of customers

### *Customer experience*

Customer experience is the most obvious concept that comes to mind in the context of this research since the research goal is concerned with the experience that customers experience. Multiple definitions of customer experience are found in literature, but they commonly acknowledge the experience of the interaction between a customer and a company. Gentile, Spiller, & Noci (2007) state that it is the experience of a customer in interacting with a product, a company, or part of its organization, to which the customer reacts. Meyer & Schwager (2007) state that it encompasses the total experience that is formed by the experience of the search, purchase, consumption, and after-sales. This is an experience of the individual and resonates on different levels (rational, emotional, sensorial, physical, and spiritual). It is the internal and subjective response a customer has when having any direct contact (e.g. purchase, use, service) or indirect contact (word-of-mouth, advertisement, news reports) with a company. Verhoef, et al. (2009) add to that by stating that customer experience is formed by aspects that the company can control (e.g. service interface, atmosphere in-store, price) but that it is also influenced by elements that are out of the company's control (e.g. opinions of others, customer's purpose).

Lemke, Clark, & Wilson (2011) studied customer experience quality, combining definitions of customer experience with definitions of the quality of a product or service. The latter is commonly defined as the perceived judgement of the customer about its excellence or superiority. This judgement is based on comparing a product or service quality against expectations. This can be reflected onto customer experience. Customers have goals, which are paired with expectations or desires for the customer experience in achieving those goals. Therefore, customers also have a perceived judgement about the actual experience. A common misunderstanding is that the customer experience is ideal whenever positive emotions are sparked. Instead, it is more related to the goals that customers have in mind. The customer experience quality is dependent of the customer's judgement of direct and indirect contributions of the organization to these goals.

### *Customer service*

In the context of the financial service industry, customer service is defined as activities that involve episodes of interactions between customers and company employees for inquiries, to request changes to a policy, or to conduct financial transactions (Ray, Muhanna, & Barney, 2005). Reflecting this definition of customer services to the concept of a QAS leads to scope customer services down to interactions between a customer and a company for inquiries (i.e. questions). Activities required to handle requests for changes to a policy or to conduct financial transactions are more complex and are not directly related to the main goal of a QAS which is answering questions. However, it is imaginable that these kinds of tasks are to be covered as well in a QAS based on a customer's request proposed to the system, but as of now this is not in scope.

Dixon, Freeman, & Toman (2010) wrote an article that emphasizes companies to stop delighting their customer wherever possible when it comes to delivering customer service. They concluded that consumers rather punish companies for delivering bad service than to reward them for delightful service, which applies to both phone-based as self-service interactions. They propose two critical findings. First, customer loyalty is not built by delighting customers. Instead, reducing the effort to get their problems solved does. Second, acting deliberately on this insight can help to improve customer services, reduce customer service costs, and decrease customer churn.

Customers resent having to contact a company repeatedly, having to repeat information, and having to switch service channels. Customer services should be all about helping a customer to solve a problem quickly and easily. Dixon, Freeman, & Toman (2010) propose five tactics to follow for low-customer-effort service. First, do not only resolve the current issue but solve the next one as well (forward-resolve) to prevent the need for repeated contact about issues related to previous ones. Second, CSRs should be able to address the emotional side of customer interactions, to prevent repeated calls due to emotional disconnect, and to tailor the responses to the customer's personality type. Third, switching between channels should be minimized by increasing self-service

"stickiness". Their research showed that 57% of inbound calls came from customers who first consulted the website, and the profusion of (self-service) channels can be overwhelming for the customer to choose a suiting channel. Fourth, customer effort should be reduced by using feedback from those customers who are disgruntled or struggling (unhappy customers). Fifth, empower the front line to deliver a low-effort experience, by valuing quality of the delivered service over speed. Assessing a CSR purely on performance by means of productivity metrics (e.g. average handle time) gets in the way of making the customer's experience easy.

### *Self-service technology*

Dixon, Freeman, & Toman (2010) address the importance of self-servicing. They state that a massive shift in customer's preferences are to be found, pointing out that customers do not overwhelmingly prefer service via live contact (by phone) to self-service anymore. Customer are in fact indifferent. A self-service is defined as the possibility for a customer to produce a service independently, which means without direct CSR involvement. The SSTs that make this possible are technological interfaces (e.g. an ATM, banking by phone, services over the Internet) (Meuter, Ostrom, Roundtree, & Bitner, 2000). In categorizing an SST, its purpose (customer service, transactions, or self-help) and its interface (telephone/interactive voice response, online/internet, interactive kiosks, or video/CD) are examined. A QAS can be plotted into this categorization, with a purpose of providing customer service and self-help and an online/internet interface.

The lack of interpersonal contact in technology-based service does not mean that companies cannot please their customers through Internet-based self-service. Customer who choose such technology may not anticipate an encounter that is warm or friendship-like and may remain loyal even without the social benefits (Yen, 2005). Yen examined attributes important for customer satisfaction with Internet-based SST. The attributes efficiency, ease of use, performance, perceived control, and convenience are identified as attributes that have a significant impact on users' satisfaction with an Internet-based SST. Three segments of consumers are identified among the users based on their scores on technology readiness drives and inhibitors (the speed of adopting technology): explorers, pioneers, and sceptics. The importance of the attributes varies per consumer type because it depends on their readiness to adopt the technology.

Meuter, Ostrom, Roundtree, & Bitner (2000) identified three major groups of factors that lead to a satisfactory evaluation of the experience when using an SST. The first category covers the ability of an SST to bail customers out of immediate troubling situations, which is especially a valuable advantage over competitors who do not offer such an SST. This requires an understanding of what types of intense or immediate needs customers experience. The second and largest category covers the relative advantages customers perceive to get by using an SST. Typically, this involves the customer's comparison to an interpersonal service delivery, and the perceived advantages are often related to time, ease of use, and access. Customers seem to believe that they are more effective at producing the service opposed to the CSR, and an SST enables the customer to avoid that and produce on their own an at their own convenience. The third category covers the novelty of the technology and the ability to perform services. SSTs need to do what is intended to do with them. Its capabilities can fascinate the customer and surprise them when services are performed successfully. Meuter, Ostrom, Roundtree, & Bitner (2000) also identified four major groups of factors that lead to a dissatisfactory evaluation of the experience when using an SST. The first and largest category covers technology failures. This includes situations in which customers are prevented from using the SST. Customer confronted with a technology failure can switch service providers, revert to interpersonal service delivery (which can overload the CSRs in the case of technology failures), or not use the service at all or at a later moment. This has effects on revenue streams and that highlights the importance of maintenance for SSTs for continues effectiveness. The second category is concerned with process failures, meaning that a breakdown occurs while using the SST before completion of the service. This is a problem since customers may assume that the process has been completed, which may cause complications when the incompletion becomes apparent. The third category covers poor design for both the technology interface and the service process. SSTs should be designed with the customer in mind and be integrated in the overall service design. The fourth category involves customer-driven failures. Some of the dissatisfying encounters with SSTs are due to the influence that the interactions of customers have on the nature of the outcome that they experience. Customers are aware of this and are willing to take blame for it but acquiring their feedback and training their SST interaction can limit these failures.

Johnson, Bardhi, & Dunn (2008) studied paradoxes that affect customer satisfaction with SST in a banking context. Performance ambiguity and trust in technology are the psychological reactions that are considered to cause an overall satisfaction outcome. Essential satisfiers for an SST are consumer need fulfillment and

enablement of freedom (added motivation to explore the technology). However, enslavement (the opposite of freedom; using SST beyond a level required by their circumstance) also has a positive effect. The study identifies fear and chaos and creation of new needs as important dissatisfying concerns for customers. Dissatisfiers increase performance ambiguity, which undermines the trust in the technology, and subsequently the customer satisfaction with SST.

Ding, Hu, & Sheng (2011) state that previous studies show that dissatisfying self-service experiences often result from designed services that do not meet customers' needs, wants, or preferences. Customers demand better control, convenience, and appropriate personal assistance for service fulfillment. Ding, Hu, & Sheng (2011) propose a conceptual framework, the e-SELFQUAL scale, to examine the quality of online self-service by assessing its perceived (cognitive) control, service convenience, customer service, and service fulfillment. Results from their research suggest that online service providers should ensure service fulfillment, enable customers to control the service process, offer considerable time and effort savings, and provide timely customer service and assistance. Lin & Hsieh (2011) also examined factors affecting SST quality by developing a model for SST quality measurement, the SSTQUAL scale. This scale considers the factors functionality, enjoyment, security/privacy, assurance, design, convenience, and customization. From their results, design appears to be the most important facet for SST quality, followed by security/privacy, assurance, and functionality.

### Human-computer interaction

Apart from designing systems that provide certain services, it is important to design interactive computer systems to be effective, efficient, easy, and enjoyable to use (Dix, Finlay, Abowd, & Beale, 2004). The HCI domain is all about how to realize these goals. The problem that occurs with human users is considering the human and contextual part of a system while other parts of the system are understood and designed with rigor. This is difficult but inescapable and has led to new fields of study of which HCI is a fairly recent field. The roots of HCI lie in established disciplines that studied the interaction between humans and machines. Ergonomics research has been focusing primarily on physical characteristics of machines and systems and their affection on user performance. Human Factors research added to that by looking at cognitive issues as well. Both disciplines are concerned with user performance in the context of any form of system. HCI research became a distinctive discipline when the use of computers became widespread and research on the interaction between people and computers got attention, concerning its physical, psychological, and theoretical aspects. HCI development was also influenced by information science and technology because of new ways of storing, accessing, and utilizing information and their influence in the workplace, requiring it to comply with requirements and constraints on the job.

HCI is a central concern in computer science and system design. It involves the design, implementation, and evaluation of interactive systems in the context of user's task work. In HCI, a user can be an individual but also a group or sequence of users, as long as it is a person or people fulfilling a task using the technology. The computer can be any kind of technology, ranging from a workstation to an embedded system, and a system may even include non-computerized parts including other people. The interaction is any communication between the user and the computer. Direct communication involves a dialog with direct feedback and control during the performance of the task, and indirect communication may involve batch processing or intelligent sensors that control the environment. The interaction with a computer is always purposed to accomplish something for the user. HCI has no unified theory, but there is an underlying principle for views on HCI: the system must support the user's task, and if the system forces the user to adopt an unacceptable mode of work then it is not usable.

# Appendix B: Screen captures of question-answering system implementations of Dutch organizations

*ING: question-screen*

## *ING: answer-screen*

## *a.s.r.: question-screen*



## *a.s.r.: answer-screen*

## *T-Mobile: question-screen*

## T-Mobile: answer-screen

## *Zalando: question-screen*

## *Zalando: answer-screen*

DAMES | HEREN | KINDEREN     ▸ zalando     Inloggen   verlanglijstje   winkelwagen

Nieuw   Kleding   Schoenen   Accessoires   Sport   Premium   Merken   OUTLET     Zoek

HULP > RETOUR & TERUGBETALING

### Hoe stuur ik een artikel retour?

Via de onderstaande opties kan je een artikel van Zalando aan ons terugsturen. Stuur je een Partner artikel terug? Volg dan dit proces.

• Geef je retourzending kosteloos af bij een PostNL punt

• Maak een afhaalopdracht met onze ophaalservice ZALANDO HAALT OP

Op het bijgeleverde retourformulier kan je de artikelen aankruisen die je terug gaat sturen. Dit formulier leg je samen met de artikelen in de doos en bovenop de doos plak je de ontvangen retoursticker. Zorg er wel voor dat de originele barcode van de doos wordt verwijderd.

**EEN ARTIKEL RETOURNEREN**

Als (een deel van) je bestelling uit partnerartikelen bestaat, dan verzoeken wij deze in het aparte pakket waarin je ze hebt ontvangen ook weer terug te sturen direct naar de partner. De documenten die je nodig hebt, om dit te doen, vind je in het pakket zelf. Het online aanmelden hiervan is niet mogelijk.

Als je de overeenkomst herroept, betalen wij het aankoopbedrag (inclusief standaard verzendkosten) zo snel mogelijk terug, maar uiterlijk binnen **14 dagen** na herroeping. Wij mogen wachten met de terugbetaling tot wij de artikelen hebben teruggekregen, of je hebt aangetoond dat je de artikelen hebt teruggezonden, al naar gelang welk tijdstip eerst valt.

Was dit antwoord nuttig?   JA   NEE

### Neem contact met ons op:

BEL ONS     MAIL ONS
AANBEVOLEN

**10% korting bij onze nieuwsbrief!**
Schrijf je in voor inspiratie en aanbiedingen
Schrijf je in

| HULP | | KORTINGSCODE & CADEAUBON | ZALANDO BUSINESS |
|---|---|---|---|
| Alle veelgestelde vragen | Betalen op rekening | Cadeaubon kopen | Corporate Website |
| Terugbetaling | Wachtwoord wijzigen | Cadeaubon inwisselen | Vacatures |
| Verzenden | Aanmelden nieuwsbrief | Kortingscode inwisselen | Zalando Marketing Services |
| Retourneren | | Voorwaarden kortingscode | Marketing aanvragen |
| | | Cadeaubonnen Voor Bedrijven | Pers |

WIJ VERSTUREN MET     BETAALWIJZEN     MAKKELIJK ONLINE SHOPPEN     VERIFIED SERVICE

Gratis levering

Gratis retour

100 dagen retourrecht

## KLM: question-screen

## KLM: answer-screen

## Appendix C: User experience questionnaire for question-answering systems

# User experience questionnaire for question-answering systems

Did the system give you the feeling that you are directing the question-answering process?

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Not at all | ○ | ○ | ○ | ○ | ○ | Extremely |

Did the system give you the feeling that you are capable and effective with your actions?

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Not at all | ○ | ○ | ○ | ○ | ○ | Extremely |

Did the system give you the feeling that you are a valued user who's input about the process influences it for you and others?

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Not at all | ○ | ○ | ○ | ○ | ○ | Extremely |

Did the system give you the feeling that you can easily get in contact with people that care about you and your situation?

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Not at all | ○ | ○ | ○ | ○ | ○ | Extremely |

Did the system give you the feeling that you are in safe hands and can prevent escalation of concerning events?

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Not at all | ○ | ○ | ○ | ○ | ○ | Extremely |

Did the system give you the feeling that you can get enjoyment and pleasure as well besides getting an answer?

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Not at all | ○ | ○ | ○ | ○ | ○ | Extremely |

Did you feel inspired?

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Not at all | ○ | ○ | ○ | ○ | ○ | Extremely |

Did you feel alert?

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Not at all | ○ | ○ | ○ | ○ | ○ | Extremely |

Did you feel excited?

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Not at all | ○ | ○ | ○ | ○ | ○ | Extremely |

Did you feel enthusiastic?

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Not at all | ○ | ○ | ○ | ○ | ○ | Extremely |

Did you feel determined?

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Not at all | ○ | ○ | ○ | ○ | ○ | Extremely |

Did you feel afraid?

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Not at all | ○ | ○ | ○ | ○ | ○ | Extremely |

Did you feel upset?

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Not at all | ○ | ○ | ○ | ○ | ○ | Extremely |

Did you feel nervous?

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Not at all | ○ | ○ | ○ | ○ | ○ | Extremely |

Did you feel scared?

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Not at all | ○ | ○ | ○ | ○ | ○ | Extremely |

Did you feel distressed?

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Not at all | ○ | ○ | ○ | ○ | ○ | Extremely |

The system was:

|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 |  |
|---|---|---|---|---|---|---|---|---|
| Confusing | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Structured |

The system was:

|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 |  |
|---|---|---|---|---|---|---|---|---|
| Impractical | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Practical |

The system was:

|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 |  |
|---|---|---|---|---|---|---|---|---|
| Unpredictable | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Predictable |

The system was:

|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 |  |
|---|---|---|---|---|---|---|---|---|
| Complicated | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Simple |

The system was:

|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 |  |
|---|---|---|---|---|---|---|---|---|
| Dull | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Captivating |

The system was:

|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 |  |
|---|---|---|---|---|---|---|---|---|
| Tacky | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Stylish |

The system was:

|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 |  |
|---|---|---|---|---|---|---|---|---|
| Cheap | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Premium |

The system was:

|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 |  |
|---|---|---|---|---|---|---|---|---|
| Unimaginative | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Creative |

The system was:

|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 |  |
|---|---|---|---|---|---|---|---|---|
| Good | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Bad |

The system was:

|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 |  |
|---|---|---|---|---|---|---|---|---|
| Beautiful | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Ugly |

To what extent did the system cause your experience?

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Very small | ○ | ○ | ○ | ○ | ○ | Very large |

Choose what is most applicable for you:

○ I gather detailed information

○ I do not gather information

Choose what is most applicable for you:

○ I ignore advice

○ I depend on advice

Did you use the system individually?

○ Yes

○ No

Where there other people in the same room/space when you used the system?

○ Yes

○ No

On what kind of device did you use the system?

○ Computer

○ Tablet

○ Smartphone

Where did you use the system?

○ Inside (for example, at home or at work)

○ Outside (for example, during a walk or on the bus)

# Appendix D: Screen captures of the generic question-answering system proof of concept

## *Question-screen*

## *Answer-screen 1: answerable user question scenario*

*Logo*

*Main website navigation pane; category x, category y, category z, etc.*

*User login*

<Answerable user question>

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum. More information can be found here.

Is this answer helpful?

◯ Yes   ◯ No

### Default FAQ

| FAQ 1 | ▼ |
|---|---|
| FAQ 2 | ▼ |
| FAQ 3 | ▼ |
| FAQ 4 | ▼ |
| FAQ n | ▼ |

### Default (self-)service topics

| Category a | Category b | Category c |
|---|---|---|
| > Topic x | > Topic x | > Topic x |
| > Topic y | > Topic y | > Topic y |
| > Topic z | > Topic z | > Topic z |

| Category d | Category e | Category f |
|---|---|---|
| > Topic x | > Topic x | > Topic x |
| > Topic y | > Topic y | > Topic y |
| > Topic z | > Topic z | > Topic z |

### Interpersonal contact options

| Option a | Option b | Option c |
|---|---|---|
| Additional information | Additional information | Additional information |
| > Get in contact | > Get in contact | > Get in contact |

*Bottom website pane; quick links, about, privacy, cookies, disclaimer, social media channels, etc.*

## *Answer-screen 1: answerable user question scenario → "Yes"*



Main website navigation pane; category x, category y, category z, etc.

User login

\<Answerable user question\>

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum. More information can be found here.

Thank you for your feedback.

**Default FAQ**

| FAQ 1 | ▼ |
| FAQ 2 | ▼ |
| FAQ 3 | ▼ |
| FAQ 4 | ▼ |
| FAQ n | ▼ |

**Default (self-)service topics**

| Category a | Category b | Category c |
|---|---|---|
| > Topic x | > Topic x | > Topic x |
| > Topic y | > Topic y | > Topic y |
| > Topic z | > Topic z | > Topic z |

| Category d | Category e | Category f |
|---|---|---|
| > Topic x | > Topic x | > Topic x |
| > Topic y | > Topic y | > Topic y |
| > Topic z | > Topic z | > Topic z |

**Interpersonal contact options**

| Option a | Option b | Option c |
|---|---|---|
| Additional information | Additional information | Additional information |
| > Get in contact | > Get in contact | > Get in contact |

Bottom website pane; quick links, about, privacy, cookies, disclaimer, social media channels, etc.

## *Answer-screen 1: answerable user question scenario → "No"*

Logo

*Main website navigation pane; category x, category y, category z, etc.*

User login

<Answerable user question>

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum. More information can be found here.

Can you tell us why?

Send

**Default FAQ**

| FAQ 1 | ▼ |
| --- | --- |
| FAQ 2 | ▼ |
| FAQ 3 | ▼ |
| FAQ 4 | ▼ |
| FAQ n | ▼ |

**Default (self-)service topics**

| Category a | Category b | Category c |
| --- | --- | --- |
| > Topic x | > Topic x | > Topic x |
| > Topic y | > Topic y | > Topic y |
| > Topic z | > Topic z | > Topic z |

| Category d | Category e | Category f |
| --- | --- | --- |
| > Topic x | > Topic x | > Topic x |
| > Topic y | > Topic y | > Topic y |
| > Topic z | > Topic z | > Topic z |

**Interpersonal contact options**

| Option a | Option b | Option c |
| --- | --- | --- |
| Additional information | Additional information | Additional information |
| > Get in contact | > Get in contact | > Get in contact |

*Bottom website pane; quick links, about, privacy, cookies, disclaimer, social media channels, etc.*

## *Answer-screen 1: answerable user question scenario → "No"*

## *Answer-screen 2: unspecific user question scenario*

*Logo*

*Main website navigation pane; category x, category y, category z, etc.*

*User login*

<Unspecific user question>

Indicate what you're looking for:
- FAQ/topic 1
- FAQ/topic 2
- FAQ/topic n

**Default FAQ**

| FAQ 1 | ▼ |
| FAQ 2 | ▼ |
| FAQ 3 | ▼ |
| FAQ 4 | ▼ |
| FAQ n | ▼ |

**Default (self-)service topics**

| Category a | Category b | Category c |
|---|---|---|
| > Topic x | > Topic x | > Topic x |
| > Topic y | > Topic y | > Topic y |
| > Topic z | > Topic z | > Topic z |

| Category d | Category e | Category f |
|---|---|---|
| > Topic x | > Topic x | > Topic x |
| > Topic y | > Topic y | > Topic y |
| > Topic z | > Topic z | > Topic z |

**Interpersonal contact options**

| Option a | Option b | Option c |
|---|---|---|
| Additional information | Additional information | Additional information |
| > Get in contact | > Get in contact | > Get in contact |

*Bottom website pane; quick links, about, privacy, cookies, disclaimer, social media channels, etc.*

## *Answer-screen 3: unanswerable user question scenario*

*Logo*

*Main website navigation pane; category x, category y, category z, etc.*

*User login*

<Unanswerable user question>

There is no answer to that question. Reformulate your question.

**Default FAQ**

| FAQ 1 | ▼ |
| FAQ 2 | ▼ |
| FAQ 3 | ▼ |
| FAQ 4 | ▼ |
| FAQ n | ▼ |

**Default (self-)service topics**

| Category a | Category b | Category c |
|---|---|---|
| > Topic x | > Topic x | > Topic x |
| > Topic y | > Topic y | > Topic y |
| > Topic z | > Topic z | > Topic z |

| Category d | Category e | Category f |
|---|---|---|
| > Topic x | > Topic x | > Topic x |
| > Topic y | > Topic y | > Topic y |
| > Topic z | > Topic z | > Topic z |

**Interpersonal contact options**

| Option a | Option b | Option c |
|---|---|---|
| Additional information | Additional information | Additional information |
| > Get in contact | > Get in contact | > Get in contact |

*Bottom website pane; quick links, about, privacy, cookies, disclaimer, social media channels, etc.*

## *Answer-screen 3: unanswerable user question scenario*

# Appendix E: Screen captures of the enriched generic question-answering system proof of concept

*Question-screen*

### Answer-screen 1: answerable user question scenario

Logo

*Main website navigation pane; category x, category y, category z, etc.*

User login

Know that you can always get in contact with a customer service representative if necessary.

| Option a ▾ | Option b ▾ | Option c ▾ |

Your question:   <Answerable user question>

Score! Your question has resulted in the following answer:

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum. More information can be found here.

Is this answer useful for someone you know? Share it!

**Share**

Is this answer helpful? Your feedback helps us to improve the quality and range of answers.

◯ Yes   ◯ No

Based on your question, you might be interested in reading the following article:
- <Title of article related to the question/answer>

Let us know what you think!

### Do you want to try another question?

Today, 250 helpful answers were found. In total, 45000 helpful answers were found.

*You can ask your question in natural language or just use keywords*   🔍

**Default FAQ** ▾

**Default (self-)service topics** ▾

*Bottom website pane; quick links, about, privacy, cookies, disclaimer, social media channels, etc.*

### Answer-screen 1: answerable user question scenario → "Share"



### Answer-screen 1: answerable user question scenario → "Share" → "Share"

### Answer-screen 1: answerable user question scenario → "Yes"

*Main website navigation pane; category x, category y, category z, etc.*

Logo

User login

Know that you can always get in contact with a customer service representative if necessary.

| Option a ▼ | Option b ▼ | Option c ▼ |

**Your question:** <Answerable user question>

Score! Your question has resulted in the following answer:

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum. More information can be found here.

Is this answer useful for someone you know? Share it!

**Share**

Thank you for your feedback. We are glad that you found yourself a helpful answer!

Let us know what you think!

Based on your question, you might be interested in reading the following article:
• <Title of article related to the question/answer>

### Do you want to try another question?

Today, 250 helpful answers were found. In total, 45000 helpful answers were found.

*You can ask your question in natural language or just use keywords* 🔍

**Default FAQ** ▼

**Default (self-)service topics** ▼

*Bottom website pane; quick links, about, privacy, cookies, disclaimer, social media channels, etc.*

## Answer-screen 1: answerable user question scenario → "No"



Logo

*Main website navigation pane; category x, category y, category z, etc.*

User login

Know that you can always get in contact with a customer service representative if necessary.

| Option a ▼ | Option b ▼ | Option c ▼ |

Let us know what you think!

Your question:    <Answerable user question>

Score! Your question has resulted in the following answer:

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum. More information can be found here.

Is this answer useful for someone you know? Share it!

**Share**

Sorry to hear that the system failed to be helpful. Can you indicate why the answer is not helpful?

☐ It is not answering the question  ☐ It is incomplete  ☐ It is unclear  ☐ It is inappropriate

*Feel free to add an explanation*

**Send**

Based on your question, you might be interested in reading the following article:
- <Title of article related to the question/answer>

### Do you want to try another question?

Today, 250 helpful answers were found. In total, 45000 helpful answers were found.

*You can ask your question in natural language or just use keywords*  🔍

**Default FAQ** ▼

**Default (self-)service topics** ▼

*Bottom website pane; quick links, about, privacy, cookies, disclaimer, social media channels, etc.*

## *Answer-screen 1: answerable user question scenario → "No" → "Send"*

Logo

*Main website navigation pane; category x, category y, category z, etc.*

User login

Know that you can always get in contact with a customer service representative if necessary.

| Option a ▼ | Option b ▼ | Option c ▼ |

Your question:   <Answerable user question>

Score! Your question has resulted in the following answer:

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum. More information can be found here.

Is this answer useful for someone you know? Share it!

**Share**

Thank you for your feedback. It enables us to improve for you and your fellow users!

Based on your question, you might be interested in reading the following article:
• <Title of article related to the question/answer>

### **Do you want to try another question?**

Today, 250 helpful answers were found. In total, 45000 helpful answers were found.

*You can ask your question in natural language or just use keywords* 🔍

**Default FAQ** ▼

**Default (self-)service topics** ▼

Let us know what you think!

*Bottom website pane; quick links, about, privacy, cookies, disclaimer, social media channels, etc.*

## *Answer-screen 2: unspecific user question scenario*

Logo

*Main website navigation pane; category x, category y, category z, etc.*

User login

Know that you can always get in contact with a customer service representative if necessary.

| Option a ▼ | Option b ▼ | Option c ▼ |

Your question:  &lt;Unspecific user question&gt;

Almost there! The system is not sure about the answer yet. Can you specify what your question is about?
- FAQ/topic 1
- FAQ/topic 2
- FAQ/topic n

### Do you want to try another question?

Today, 250 helpful answers were found. In total, 45000 helpful answers were found.

*You can ask your question in natural language or just use keywords* 🔍

**Default FAQ** ▼

**Default (self-)service topics** ▼

Let us know what you think!

*Bottom website pane; quick links, about, privacy, cookies, disclaimer, social media channels, etc.*

## *Answer-screen 3: unanswerable user question scenario*

Logo

*Main website navigation pane; category x, category y, category z, etc.*

User login

Your question:  &lt;Unanswerable user question&gt;

Sorry! The system unfortunately could not answer your question. This might be caused by the lack of an answer on our side or a question that is too complex for the system to handle automatically. You can always try to ask another question but keep in mind that a customer service representative is always available to help you personally. Please know that any problem can be solved, possibly by yourself but definitely with us: solving problems starts here!

| Option a | Option b | Option c |
| Additional information | Additional information | Additional information |
| > Get in contact | > Get in contact | > Get in contact |

### Do you want to try another question?

Today, 250 helpful answers were found. In total, 45000 helpful answers were found.
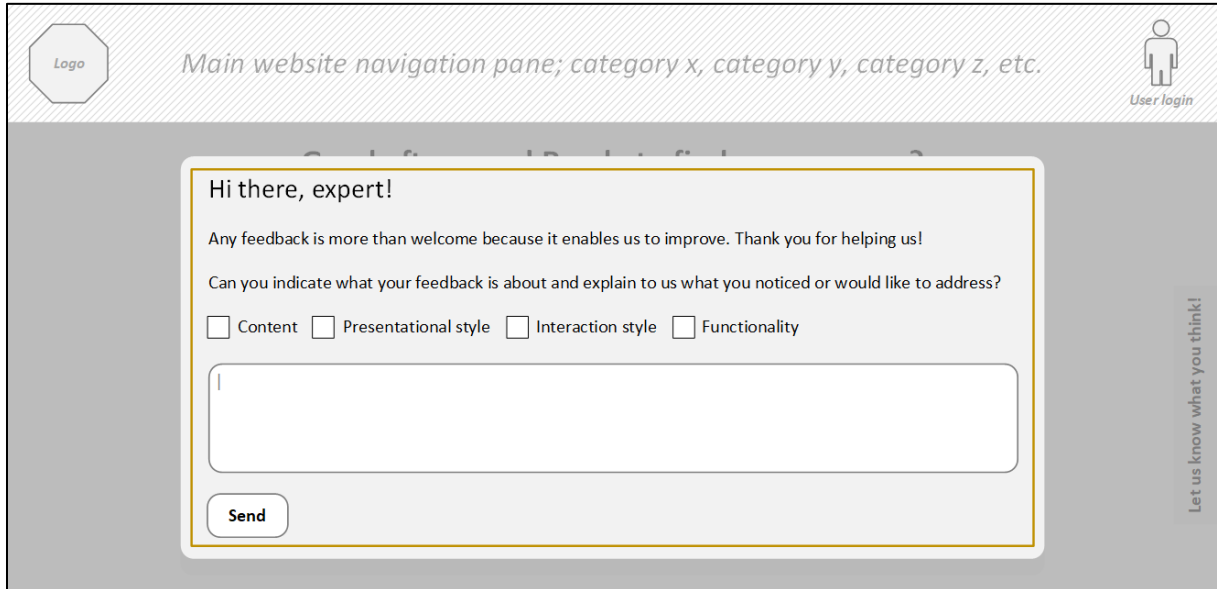
*You can ask your question in natural language or just use keywords* 🔍

**Default FAQ** ▼

**Default (self-)service topics** ▼

Let us know what you think!

*Bottom website pane; quick links, about, privacy, cookies, disclaimer, social media channels, etc.*

## *Feedback-popup*

Main website navigation pane; category x, category y, category z, etc.

Logo

User login

Let us know what you think!
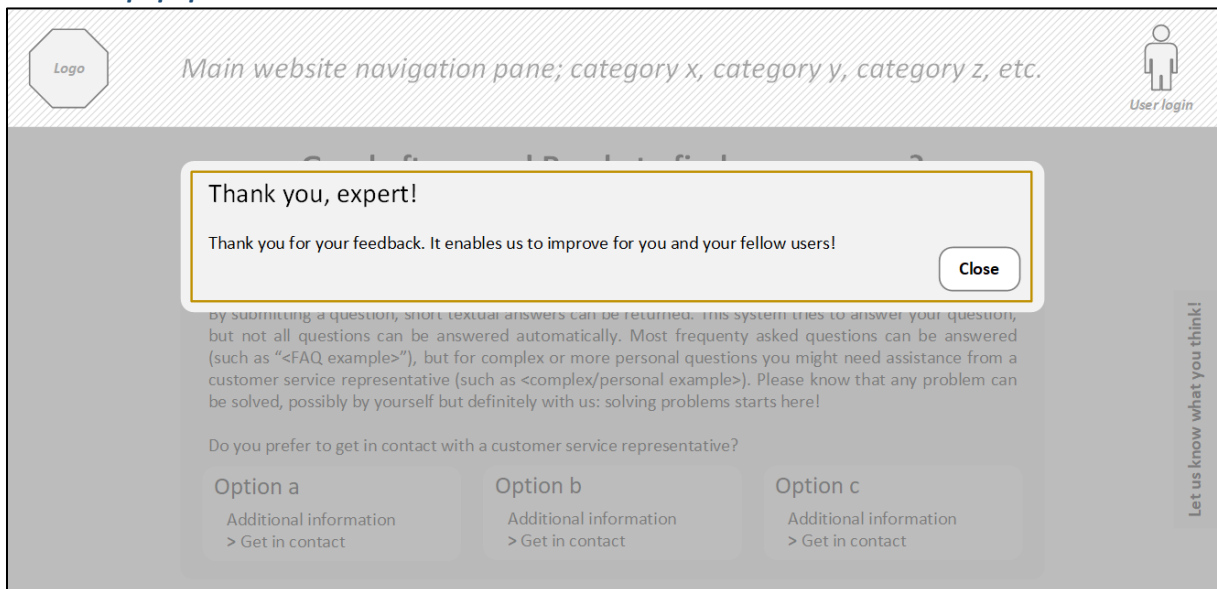
**Hi there, expert!**

Any feedback is more than welcome because it enables us to improve. Thank you for helping us!

Can you indicate what your feedback is about and explain to us what you noticed or would like to address?

☐ Content ☐ Presentational style ☐ Interaction style ☐ Functionality

Send

## *Feedback-popup → "Send"*

Main website navigation pane; category x, category y, category z, etc.

Logo

User login

Let us know what you think!

**Thank you, expert!**

Thank you for your feedback. It enables us to improve for you and your fellow users!

Close

By submitting a question, short textual answers can be returned. This system tries to answer your question, but not all questions can be answered automatically. Most frequenty asked questions can be answered (such as "<FAQ example>"), but for complex or more personal questions you might need assistance from a customer service representative (such as <complex/personal example>). Please know that any problem can be solved, possibly by yourself but definitely with us: solving problems starts here!

Do you prefer to get in contact with a customer service representative?

Option a
Additional information
> Get in contact

Option b
Additional information
> Get in contact

Option c
Additional information
> Get in contact

## User experience questionnaire-popup

Main website navigation pane; category x, category y, category z, etc.

*Logo*

*User login*

**Hi there, expert!**

You used the system a few times. Nice! Do you want to fill in a questionnaire regarding your experience with the system? It helps us to improve the experience of the system for you and your fellow users!

**Yes**  **No**

Score! Your question has resulted in the following answer:

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum. More information can be found here.

Is this answer useful for someone you know? Share it!

Share

Let us know what you think!