A NOVEL HYBRID KNOWLEDGE RETRIEVAL APPROACH FOR ONLINE CUSTOMER SERVICE PLATFORMS

Research paper

Bedué, Patrick, University of Ulm, Ulm, Germany, patrick.bedue@uni-ulm.de

Graef, Roland, University of Ulm, Ulm, Germany, roland.graef@uni-ulm.de

Klier, Mathias, University of Ulm, Ulm, Germany, mathias.klier@uni-ulm.de

Zolitschka, Jan Felix, University of Ulm, Ulm, Germany, jan.zolitschka@uni-ulm.de

Abstract

The rising number of internet users has considerably contributed to the need of online customer services. Against this background, many organisations aim to tap the potential of online customer service platforms where employees answer customer questions within a short time frame. However, these platforms also go along with several challenges for organisations. Indeed, with their rising popularity, the knowledge contained in online customer service platforms as well as the number of (potentially redundant) customer questions, are more and more increasing. Ensuring a high quality and consistent online customer service, organisations face the challenge to manage their knowledge in an efficient way and require a fast and effective knowledge retrieval. In this context, assisted systems for employees can be an appropriate means. However, there is still a lack of integrated knowledge retrieval approaches, considering both structured and textual information, which are both of major importance in the context of online customer service platforms. To address this problem, we propose a new hybrid knowledge retrieval approach combining structured and textual information in an integrated perspective. We demonstrate and evaluate our approach using the case of a major German insurer where the approach could be successfully applied.

Keywords: Knowledge Retrieval, Online Customer Service Platform, Case-based Reasoning.

1 Introduction

In recent years, the number of internet users has risen from 1,024 million in 2005 up to 3,578 million in 2017 (Statista, 2017). This increase has considerably contributed to the need of online customer services. Indeed, it has become a trend in the today's fast-moving world to have quick and easy access to information via the Internet (Decker, 2016). As a consequence, it is promising for organisations to operate online customer service platforms, where customer questions are answered by employees within a short time frame (Heidelberger and Sarnikar, 2008; Zhang et al., 2011). Hewlett Packard (HP), for instance, as multinational information technology company successfully operates an online customer service platform[1] with over 1,000,000 topics to support their customer service. In this platform customers get easy access to the free-text descriptions (*textual information*) of many previous customer problems and respective solutions organised according to hierarchical topics (e.g. "Notebooks") and subtopics (e.g. "Notebook Video, Display and Touch") (*structured information*).

In practice, on the one hand many organisations already aim to tap the potential of online customer service platforms as a new way to improve customer service. On the other hand, the operation of online customer service platforms goes along with several new challenges for organisations. First, organisations lose more than $62 billion on average every year due to poor online customer service mainly driven by customers abandoning business processes (Noble, 2016). Second, customers dissatisfied by the online customer service discourage further potential customers by negative word of mouth from using the online customer service platform (eMarketer, 2016; Noble, 2016; Zendesk, 2017). According to a

[1] https://h30434.www3.hp.com/

recent study by Forrester (2016), 53% of US adults reject purchasing, if the response to requested expert knowledge is unreliable, inconsistent or not delivered in time. However, currently most users of online customer service have to wait on average ten hours to get a response from an organisation, although they are only willing to wait for a maximum of four hours (Sprout Social, 2016). A recent study by McKinsey & Company (2016) even states that 75% of online customers expect service within 5 minutes.

Consequently, organisations face the challenge to manage their knowledge in an efficient way and foster the development of a fast and easy knowledge retrieval to be able to provide solutions for customer problems in time. However, the knowledge contained in online customer service platforms as well as the number of (potentially redundant) customer problems are more and more increasing (Prince, 2017). Hence, service employees are overstrained by looking for specific solutions in the vast amount of knowledge, in which searching for appropriate information seems to be like looking for a needle in a haystack. Trying to cope with this problem by hiring further employees may only be helpful in a short run. Indeed, this would cause a decentralised distribution and inconsistent levels of knowledge (Grant, 1996; Moreo et al., 2012) which hinders a high quality and consistent customer service in the long run.

To address these challenges and tap the potential of online customer service platforms, assisted systems for employees in online customer service can be used to retrieve helpful knowledge from already solved customer problems. This knowledge can support employees to provide customers with well-founded and consistent solutions regarding their problems in time. According to a study by the Oracle Corporation (2011), assisted systems promise huge potential for efficient and fast knowledge retrieval to handle the tremendous amount of customer problems. The basic idea of assisted systems for online customer service platforms is to identify and provide the service employees with similar and already solved customer problems, which can then be reused to solve new customer problems. The concept behind is that similar problems imply similar solutions. A promising approach to identify similar customer problems and respective solutions, in the following referred to as similar cases, is Case-based Reasoning (CBR). Indeed, CBR has already been successfully applied in many real-world scenarios in the context of customer support systems (Acorn and Walden, 1992; Heras et al., 2009; Lee et al., 2015; Lenz et al., 1998).

In online customer service platforms, customer problems comprise structured as well as textual information, which has to be taken into account to retrieve relevant knowledge. Actually, literature already provides well-founded CBR approaches for structured information (Burke et al., 2000; Heras et al., 2009; Ricci and Senter, 1998) and textual information (Balakrishnan et al., 2016; Burke et al., 1997; Lenz et al., 1998; Weber et al., 2005; Weis, 2015), respectively. However, there is still a lack of integrated approaches considering both types of information. To address this research gap, we propose a new hybrid knowledge retrieval approach combining structured and textual information in an integrated perspective to adequately assist employees in online customer service platforms.

Following a design-oriented approach (cf., e.g. Peffers et al., 2007), the remainder of this paper is structured as follows: In the next section, we provide an overview of the problem context and related work. In Section 3, we propose a hybrid knowledge retrieval approach combining structured and textual information for online customer service platforms. In Section 4, we demonstrate and evaluate our approach using the case of a German insurer where the approach could be successfully applied. Finally, we conclude with a summary of the findings, a discussion of limitations and an outlook on future research.

2 Problem Context and Related Work

2.1 Problem context

In online customer service platforms, employees solve customer problems. These platforms are typically built upon a hierarchical structure, for example by certain topics or subtopics, to support knowledge access. Online customer service platforms exist in various domains, for example in the areas of information technology (Dell), telecommunication (AT&T) and insurance (AMPM).[2] The following example

[2] Dell: http://en.community.dell.com/; AT&T: https://forums.att.com/; AMPM: http://www.ampminsure.org/community/

illustrates our problem context: Consider the AMPM insurance online customer service platform with a set of already solved problems, a customer having an issue concerning a disability insurance and an employee of the insurer.

Domain Categories *(Structured)*	*Disability Insurance*	*Payment*
Customer Question *(Textual)*	*Hello Experts, I have a disability insurance and recently this accident happened: … Does my occupational disability insurance pay? … Sincerely, Your Customer*	
Solution *(Textual)*	*Dear Customer, in your specific situation the occupational disability insurance will … With best regards, Your Expert*	

Figure 1: Case Structure and Illustrative Example.

The customer can describe the issue in free-text (*customer question*) providing domain-specific information (e.g., description of the disability insurance) and general information (e.g., description of the circumstances). Furthermore, the customer adds suitable structured information (*domain categories*) such as “Disability Insurance” and “Payment” with respect to the hierarchy of the underlying platform. Based on this customer problem (cf., first two rows in Figure 1) and existing knowledge, the service employee has to provide a *solution*, which is reliable and consistent (cf., last row in Figure 1). To meet these requirements, the employee has to be aware of the existing knowledge in the online customer service platform represented by the linked information for a case regarding domain categories, the customer question and the respective solution (cf., Figure 1). The AMPM platform, for example, contains more than 100,000 discussions including about 5,000 discussions on the topic “Disability Insurance”. The consolidation of all cases, in the following referred to as *case base*, constitutes the cornerstone of the knowledge in the online customer service platform. Obviously, with the increasing number of new customer problems and cases in the case base, a consistent and high quality knowledge retrieval in a short time frame becomes a highly difficult task. Therefore, approaches for online customer service platforms are needed, which are able to not only retrieve information related to a small amount of search words (cf., e.g. research area of information retrieval (Campos et al., 2015; Carpineto and Romano, 2012)) or to answer questions by combining knowledge in a constructed answer (cf., e.g. research area of question answering (Höffner et al., 2017; Kolomiyets and Moens, 2011)). In contrast, an approach is required for retrieving existing knowledge containing both, similar customer questions and related employee answers in terms of existing cases, while at the same time offering a mechanism for evolving with the number of cases solved (Heras et al., 2009; Lenz et al., 1998; Lenz et al., 1999)).

2.2 Related work and research gap

Research already provides approaches for knowledge management systems supporting employees to work more efficiently (Alavi and Leidner, 1999; Alavi and Leidner, 2001; Barão et al., 2017; Martin et al., 2017; Rubenstein-Montano et al., 2001; Lee et al., 2015; Tiwana, 2000). In the context of assisted systems for customer service, there also exist some promising approaches for automated knowledge retrieval (Acorn and Walden, 1992; Göker and Roth-Berghofer, 1999; Heras et al., 2009; Kriegsman and Barletta, 1993; Lenz and Burkhard, 1997; Lenz et al., 1999; Simoudis, 1992). To ensure customer problems are solved correctly, consistently and in time, some authors (Acorn and Walden, 1992; Kriegsman and Barletta, 1993; Simoudis, 1992) developed approaches for supporting customer service employees in the IT industry. Other assisted systems can be found in the car industry (Göker and Roth-Berghofer, 1999), the telecommunication market (Lenz and Burkhard, 1997) or in the area of automation systems (Lenz et al., 1999). As there exists a wide range of domain-specific assisted systems, Heras et al. (2009) provide an approach for a multi-domain module, which can be integrated into any organisation to support customer service.

Regardless of the domain, all mentioned approaches for automated knowledge retrieval in customer service are based on CBR. CBR is a methodology in artificial intelligence for solving problems through reusing solutions of previously solved similar cases (Aamodt and Plaza, 1994; Bergmann, 2002; de Mantaras et al., 2005; El-Sappagh and Elmogy, 2015; Martin and Plaza, 2004; Watson, 1999; Yan et al., 2014). With respect to the knowledge retrieval in online customer service platforms, CBR presents a promising approach to ensure consistent and correct solutions for customer problems by retrieving respective knowledge from the huge amount of cases already contained in the case base. Moreover, CBR is capable of retrieving relevant knowledge for the increasing number of customer problems in customer service platforms with high quality as for every solved customer problem the respective case is added to the case base. Therefore, CBR constitutes a self-learning approach, which evolves with the number of customer problems solved. In the context of online customer service platforms, customer problems comprise both, structured information (e.g., domain categories) as well as textual information (e.g., customer question). In literature the different types of CBR approaches can particularly be classified into structural CBR (Aamodt, 1991; Burke et al., 2000; Heras et al., 2009; Plaza, 1995; Ricci and Senter, 1998; Yokoyama, 1990) and textual CBR (El-Sappagh and Elmogy, 2015; Lenz et al., 1998; Weber et al., 2005; Wilson and Bradshaw, 1999). Due to the structured as well as textual information in online customer service platforms, both CBR research streams seem promising to cope with our problem.

The aim of structural CBR is to retrieve similar cases based on a set of structured attributes (Aamodt 1991; Burke et al., 2000; Heras et al., 2009; Plaza, 1995; Ricci and Senter, 1998; Yokoyama, 1990). Thereby, structured cases are often organised as flat attribute-value pairs (Behbahani et al., 2012; Bergmann, 2002; Chen and Chen, 2011; Guo et al., 2012), in an object-oriented manner (Bergmann, 2002; Bergmann and Stahl, 1998; El-Sappagh and Elmogy, 2015; Yokoyama, 1990), or in a graph or rather tree structure (Bergmann and Wilke, 1996; Heras et al., 2009; Kriegsman and Barletta, 1993; Ricci and Senter, 1998; Watson and Perera, 1998). For instance, Heras et al. (2009) built a hierarchical tree referred to as Typification Tree for indexing cases. More precisely, all cases in the case base are organised by their structured information within the Typification Tree. Then, for example regarding a customer problem with the domain categories “Disability Insurance” and “Payment”, similar cases are retrieved by searching the tree structure with respect to the particular topic and subtopic.

In context of textual CBR, literature focuses on the retrieval of similar cases based on free-text in a full- or semi-automated way (Ashley, 1991; Burke et al., 1997; Balakrishnan et al., 2016; Daniels and Rissland, 1997; Lenz et al., 1998; Sizov et al., 2015; Weber et al., 2005; Weis, 2015). For full-automated retrieval most approaches make use of well-known methods from information retrieval (Burke et al., 1997; Hammond et al., 1995; Kunze and Hübner, 1998; Lenz and Burkhard, 1997; Lenz et al., 1998; Lenz et al., 1999; Shekhar et al., 2014; Wilson and Bradshaw, 1999). For instance, Burke et al. (1997) rely on the Vector Space Model (VSM) (Salton et al., 1975) for building their FAQ Finder which retrieves the most similar frequently asked questions and the corresponding answers from the case base. Others (Kunze and Hübner, 1998; Lenz and Burkhard, 1997; Lenz et al., 1999; Shekhar et al., 2014) base their approach on the Inference Network Model (Turtle and Croft, 1989) by embedding cases into a network linked with Information Entities representing statistically identified or domain-specific phrases or terms. Links between Information Entities are built by measuring the semantic similarity between these phrases or terms. For instance, the term “disabled” and “invalid” have a close semantic meaning. Further, links from a case to Information Entities are built by the relevance of Information Entities to the specific case, for example, whether the Information Entity is contained in the case or not. By querying the network, cases can be retrieved based on the similarities between their Information Entities. In the context of CBR, this is referred to as Case Retrieval Network (Lenz and Burkhard, 1996; Lenz and Burkhard, 1997; Lenz et al., 1998; Lenz et al., 1999; Shekhar et al., 2014). Further studies provide semi-automated textual CBR approaches based on own formal languages to transform textual documents into a structured form (Aleven, 1997; Aleven and Ashley, 1997; Ashley, 1991; Brüninghaus and Ashley, 1997; Brüninghaus and Ashley, 1999; Brüninghaus and Ashley, 2005; Daniels and Rissland, 1997; Shimazu, 1998). A major drawback of these approaches is the excessive manual effort required to transform the raw text into the respective language. With regard to the vast amount of cases in online customer service platforms, semi-automated textual CBR approaches do not seem appropriate.

To sum up, CBR seems a very promising means to cope with current challenges in the context of online customer service platforms. Since a customer problem comprises structured (*domain categories*) as well as textual information (*customer question*), the approaches from structural and textual CBR are both relevant to retrieve knowledge from an online customer service platform. However, first promising CBR approaches considering structural and textual aspects (Kriegsman and Barletta, 1993; Recio et al., 2005) are not able to adequately address our problem context. Recio et al. (2005) use information extraction techniques to gain structured attributes from free-text, but only add up the results for structured and textual information, respectively, they do not take an integrated perspective from a methodical point of view. Kriegsman and Barletta (1993) primarily focus on structural CBR using a hierarchical tree but also consider simple structured features extracted from short textual descriptions. Methods from textual CBR, however, are not applied. To the best of our knowledge, so far none of the studies in CBR has considered structured and textual information in conjunction with each other while at the same time taking an integrated perspective by not only adding structured to textual information but rather combining the research streams by embedding structural CBR methods into textual CBR. To address this gap, following a design oriented approach (cf., e.g. Peffers et al., 2007), we aim at developing a novel hybrid knowledge retrieval approach combining structural textual CBR in a well-founded way.

3 Hybrid Approach for Knowledge Retrieval in Online Customer Service Platforms

3.1 Basic idea and overview of the hybrid approach

With the goal of developing an approach for automated problem solving, Aamodt and Plaza (1994) were the first to introduce the CBR Cycle consisting of the four phases Retrieve, Reuse, Revise and Retain. Following the CBR Cycle, the solution(s) for a new incoming user problem are identified by first, *retrieving* the most similar case(s) from the case base; second, *reusing* the retrieved case(s) with optional adaptations; third, *revising* the solution(s) on correctness; and fourth, *retaining* the solved case by adding it to the case base. As outlined in Section 2.1, the main challenge for organisations is the increasing number of customer problems and the huge amount of knowledge contained in the case base. Therefore, we focus in a first step on the retrieval of helpful knowledge (cf. “Retrieve phase” in Figure 2).

For the Retrieve phase all cases from the online customer service platform are available in a case base (Burke et al., 1997; Cunningham et al., 2004; Lenz et al., 1998). Following the CBR Cycle, retrieval aims at quantifying the degree of resemblance (similarity) between the customer problem and the existing cases in the case base (Liao et al., 1998). Thus, the incoming customer problem from the online customer service platform has to be processed by our hybrid approach regarding its structured as well as its textual information. Against this backdrop, we combine methods for structural knowledge retrieval from CBR with those for textual knowledge retrieval while at the same time taking an integrated perspective by not only adding structured to textual information but rather combining the respective methods (cf. “Structural knowledge retrieval” and “Textual knowledge retrieval” in Figure 2).

In the first step of our approach, the structured information of the customer problem (domain categories) is used to shrink the case base to a subset of relevant cases. To do so, we rely on approaches from structural CBR using hierarchical trees for case retrieval (Heras et al., 2009; Kriegsman and Barletta, 1993; Ricci and Senter, 1998; Watson and Perera, 1998). Thereby, we select all cases from the case base with the same domain categories compared to the customer problem (cf. “Step 1” in Figure 2). By this means, the enormous amount of cases in the case base is shrunken, which enables a more detailed analysis and an efficient textual knowledge retrieval in the next step. Here, the textual information of the customer problem (customer question) is considered for a more detailed analysis. A customer question can be highly complex in certain domains and contains on the one hand domain-specific information (e.g., description of the customer’s disability insurance) and on the other hand general information (e.g., description of the present circumstances). As the domain-specific information is the main characteristic property of a case in online customer service platforms, it is more decisive to distinguish between cases than general information. Nevertheless, general information could be necessary to distinguish between slightly different cases with similar domain-specific information, but different general information. Hence, in textual knowledge retrieval both, domain-specific (cf. “Step 2” in Figure 2) and general information are considered to determine an overall similarity (cf. “Step 3” in Figure 2). The second step of our approach builds upon the shrunken case base of the first step. To take the importance of domain-specific information into account, we reward the degree of similar domain-specific information in the customer question and the existing cases by a domain-specific weighting based on the Case Retrieval Network approach (Lenz and Burkhard, 1997; Lenz et al., 1999; Lenz et al., 1998; Shekhar et al., 2014). The third step of our approach uses the entire textual information contained in the subset of customer questions from the first step. Hence, general and domain-specific information are taken into account, whereby the importance of domain-specific information is increased by integrating the identified weights of the second step. As a result, the third step determines an overall similarity between the customer problem and each existing case in the subset of cases from the first step. Finally, the most similar cases in the case base are *retrieved* by our novel and hybrid knowledge retrieval approach based on the first and second step referring to textual CBR approaches inspired by the VSM (Burke et al., 1997; Hammond et al., 1995). In the following phases of the CBR cycle, the employee *reuses* the knowledge contained in the most similar cases to provide a correct, reliable as well as consistent solution for the customer problem. Subsequently, if the solution is *revised* on correctness, the corresponding case, comprising the customer problem and the provided solution, is *retained* in the case base. In the following subsections, we present our hybrid three-step-approach for the Retrieve phase in detail.

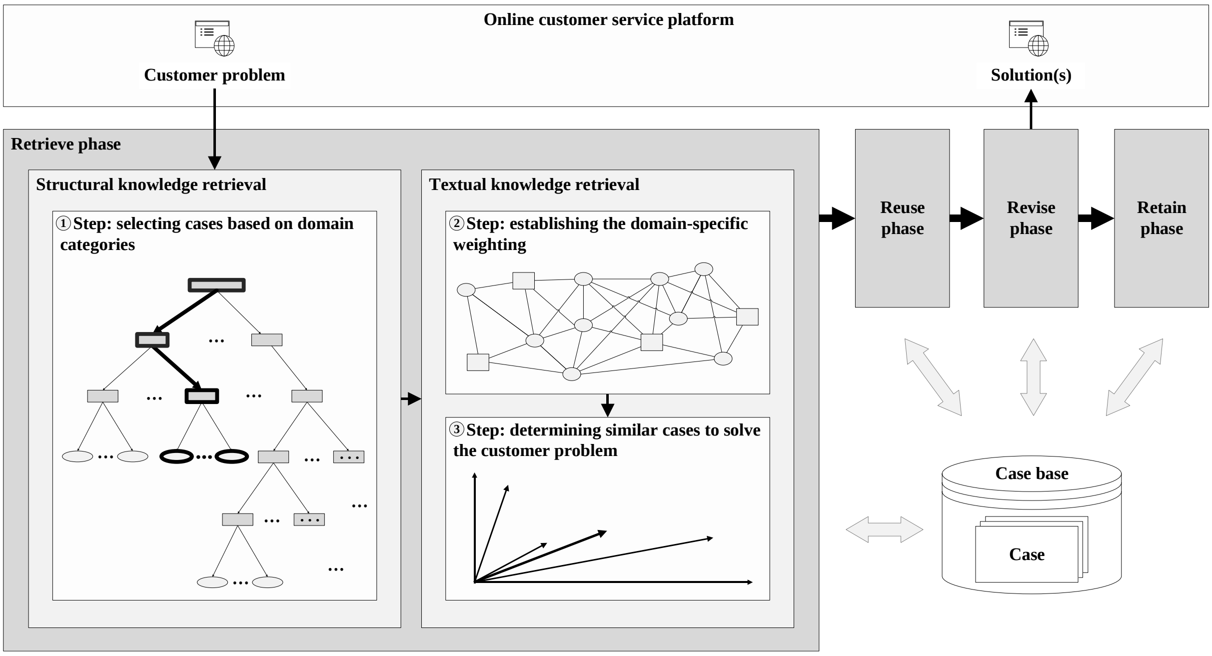

Figure 2: Hybrid knowledge retrieval approach for online customer service platforms.

3.2 First step: selecting cases based on domain categories

The aim of this step is to select cases with the same domain categories as the customer problem from the hierarchically organised cases in the case base. In case of missing domain categories for some customer problems or cases, machine learning approaches (e.g. text-to-category classification) can be applied to deduce domain categories (Cortes and Vapnik, 1995; Park et al., 1991). The domain categories can be used to shrink the case base to a smaller set of relevant cases. For example, the customer problem in Figure 1 is related to the domain categories *“Disability Insurance”* and *“Payment”*. Thus, only existing cases within these domain categories should be taken into account when searching for possible solutions. In CBR literature, hierarchical trees are often applied with convincing results for organised structured information in case retrieval contexts (Heras et al., 2009; Kriegsman and Barletta, 1993; Ricci and Senter, 1998; Watson and Perera, 1998). Consequently, we use a domain-specific Typification Tree based on the hierarchical structure of the cases in the case base to perform the structured knowledge retrieval and to shrink the case base. The shrunken case base is used as a basis for the textual knowledge retrieval in the following steps of our approach.

More precisely, the structured knowledge retrieval uses the hierarchically organised tree of all cases in the case base. Therein, each case is assigned based on its domain categories and related to the hierarchy of the underlying platform. In order to identify those cases, which are similar with respect to the customer problem P, the Typification Tree algorithm starts with the root of the tree and traverses incrementally through the nodes (Heras et al., 2009). If a node matches a domain category c_i from the set of domain categories P_C contained in the customer problem P, the child node(s) are examined. If there exist no further corresponding child node(s), the cases of the remaining subtree(s) are added to the result set $T(P_C)$. For example, regarding a customer problem containing the domain categories "Disability Insurance" and "Payment", the Typification Tree algorithm starts with the root node and subsequently examines the nodes representing "Disability Insurance" and "Payment" (cf. bold nodes in the Typification Tree in Figure 2). As a result, all cases of the subtree "Payment" are added to the result set of the tree $T(P_C)$ (cf. bold leaves in the Typification Tree in Figure 2). As customer questions could be highly complex in certain domains and address different issues, the domain categories reflect this aspect (e.g. "Payment" and "Deadline"). Thus, if different issues are addressed, the cases of multiple subtrees are added to the result set $T(P_C)$ (e.g. subtrees corresponding to the categories "Payment" and "Deadline").

To sum up, structural knowledge retrieval provides a result set of similar cases $T(P_C)$ regarding the set of domain categories P_C contained in the customer problem. The more domain categories are available in the online customer service platform and the more precisely the customer describes the problem by means of domain categories, the better the case base will be shrunken. Thus, the shrunken case base enables a more detailed analysis and efficient textual knowledge retrieval in the following steps.

3.3 Second step: establishing the domain-specific weighting

The aim of this step is to consider the domain-specific information contained in the customer question P_Q of the customer problem P, which constitutes the main characteristic property of a case. Indeed, the degree of corresponding domain-specific information can reflect the similarity of a customer problem and a specific case, concerning the domain-specific content. To bear the importance of domain-specific information in online customer service platforms in mind, a domain-specific weighting for each pair of customer question P_Q and customer question B_Q of case B is used. The domain-specific weighting can be considered as a kind of reward for the overall similarity determination. If the domain-specific information of the customer problem and the compared case is the same or nearly the same, the reward should be high et vice versa. According to CBR literature, the domain-specific information of the customer questions can be represented as Information Entities in a Case Retrieval Network, which is often applied in the context of customer service (Lenz and Burkhard, 1997; Lenz et al., 1999; Lenz et al., 1998). To derive a domain-specific weighting, we use a network similarity measure (Lenz and Burkhard, 1997; Lenz et al., 1999; Lenz et al., 1998) based on the domain-specific Information Entities of the customer questions and the Case Retrieval Network. Thus, the network similarity can be used as domain-specific weighting in the overall similarity determination in the third step.

More precisely, the textual knowledge retrieval uses a Case Retrieval Network containing all domain-specific information (Information Entities) and cases of the shrunken case base $T(P_C)$. Each Information Entity is connected to other Information Entities through similarity arcs (Lenz and Burkhard, 1996). These arcs describe the semantic similarity $sim_S(e_i, e_j)$ between two Information Entities e_i and e_j (e.g., "disabled" and "invalid" have a close semantic meaning). In addition, relevance arcs represent the relevance of an Information Entity with regard to a case (e.g., "invalid" has a high relevance regarding a case with the customer question "… and recently this accident happened: …"). To provide a weight as a reward for similar domain-specific information, all Information Entities $e_i \in IE(P_Q)$ of the customer problem P and all Information Entities $e_j \in IE(B_Q)$ of the case B are considered. If P and B share domain-specific information, their terms are connected in the Case Retrieval Network. If all domain-specific information of the customer problem is contained within the domain-specific information of the case, the network similarity measure $sim_N(P_Q, B_Q)$ yields one (cf. Equation (1)). In order to do so, we first determine whether a particular domain-specific information of the customer problem is contained in a case. Thus, we select the maximum semantic similarity $sim_S(e_i, e_j)$ between an Information Entity of the customer problem and all Information Entities of the case (cf. max function in Equation (1)). Further, regarding all Information Entities of the customer problem, the sum of the maximum similarities from all Information Entities of the customer problem $IE(P_Q)$ is determined. Finally and to gain the network similarity $sim_N(P_Q, B_Q)$, we normalize the resulting summed semantic similarities with the number of Information Entities contained in the customer problem (cf. Equation (1)).

$$sim_N(P_Q, B_Q) = \frac{1}{|IE(P_Q)|} * \sum_{e_i \in IE(P_Q)} \left(\max_{e_j \in IE(B_Q)} sim_S(e_i, e_j) \right) \tag{1}$$

To sum up, the second step provides the network similarity $sim_N(P_Q, B_Q)$ based on domain-specific information for each pair of customer questions P_Q and B_Q on the result set $T(P_C)$ of the first step. The network similarity $sim_N(P_Q, B_Q)$ is used as a domain-specific weighting in the next step.

3.4 Third step: determining similar cases to solve the customer problem

The aim of this step is to retrieve the most relevant cases regarding the customer problem so that an employee can provide a correct, reliable as well as consistent solution based on these cases. Thereby, the most similar cases of all cases $T(P_C)$ in the shrunken case base in comparison to the customer problem are determined. Further, the entire textual information of the customer question P_Q is considered to determine the overall similarities between the customer problem and each case in the shrunken case base $T(P_C)$ based on the VSM (Salton et al., 1975). In CBR literature, the VSM is often applied with convincing results for using textual information and weightings in retrieval contexts (Burke et al., 1997; Liao et al., 1998; Recio et al., 2005; Salton et al., 1975; Wilson and Bradshaw, 1999). Within the VSM, information from text is represented as a vector in a n-dimensional vector space, whereby n indicates the number of distinct words (terms). Each vector is weighted based on the relevance of the individual terms by applying well known frequency measures (Sebastiani, 2002). Afterwards, the similarity is measured by comparing vectors. Thus, we determine the overall similarity by comparing the corresponding vector of the customer problem with vectors representing the cases in the case base. Using the VSM has the advantage to include the shrunken case base $T(P_C)$ created in the first step and it allows an integration of the domain-specific weighting into the VSM weighting. Consequently, we combine the results of the first and second step with the VSM to determine an overall similarity (cf. Equation (4)).

More precisely, the VSM case retrieval uses the vector space representation of all cases in the case base and applies a similarity measure to retrieve the overall similarity between the customer questions P_Q and B_Q. The vector space representation of the customer questions P_Q and B_Q is gained by transforming each term t_i contained in these questions with respect to a chosen term weighting function $VSM_{T(P_C)}()$, e.g. $tf.idf$ (Hua et al., 2009; Salton and Buckley, 1988; Sebastiani, 2002). This results in the corresponding vector entries v_p^i and v_b^i. By performing this for each distinct term t_i contained in the n-dimensional vector space, P_Q and B_Q are transformed into appropriate weighted vectors v_p and v_b.

$$v_p^i = VSM_{T(P_C)}(t_i, P_Q), v_b^i = VSM_{T(P_C)}(t_i, B_Q) \ \forall \ i = 1, \dots, n \tag{2}$$

To integrate the domain-specific weighting into the VSM, the value of the terms represented in the vectors v_p and v_b are transformed (cf. Equation (3)). This results in the modified vectors w_p and w_b, where the corresponding vector entries w_p^i and w_b^i are modified based on the determined network similarity $sim_N(P_Q, B_Q)$ from the second step. In detail, only vector entries of terms having corresponding Information Entities are increased depending on the result of the indicator function $\chi_{t_i \in IE}$. Particularly, the domain-specific weighting from the second step is greater than zero, if the term represented by t_i belongs to the set of Information Entities IE.

$$w_p^i(P_Q, B_Q) = v_p^i * \left(1 + sim_N(P_Q, B_Q) * \chi_{t_i \in IE}\right), w_b^i(P_Q, B_Q) = v_b^i * \left(1 + sim_N(P_Q, B_Q) * \chi_{t_i \in IE}\right) \forall\, i = 1, \dots, n \quad (3)$$

Based on the modified vectors $w_p(P_Q, B_Q)$ and $w_b(P_Q, B_Q)$ a similarity measure $sim_{VSM}(\,)$ (Salton and McGill, 1983) is applied to calculate the overall similarity $SIM(P, B)$ between customer problem and each case (cf. Equation (4)). An appropriate way to determine the overall similarity is applying well-established similarity measures or machine learning approaches (Burke et al., 1997; Hammond et al., 1995; Lenz and Burkhard, 1997; Stahl, 2005; Stahl and Gabel, 2006).

$$SIM(P, B) = sim_{VSM}\left(w_b(P_Q, B_Q), w_q(P_Q, B_Q)\right) \forall\, B \in T(P_C) \quad (4)$$

To sum up, the third step represents the customer problem and all cases from the shrunken case base $T(P_C)$ as vectors in the VSM. To do so, a term weighting function $VSM_{T(P_C)}(\,)$ is applied to the customer questions P_Q and B_Q yielding in the vectors v_p and v_b. Afterwards, we apply the domain-specific weighting of the second step to obtain the modified vectors $w_p(P_Q, B_Q)$ and $w_b(P_Q, B_Q)$. As a last sub-step, the overall similarity $SIM(P, B)$ between the customer problem P and a case B is determined by using the similarity measure $sim_{VSM}(\,)$. Finally, the three-step-approach yields in a ranking of cases based on the degree of overall similarity to the customer problem. Therefore, an employee of an online customer service platform is provided with helpful knowledge in order to solve the customer problem.

4 Demonstration and Evaluation

4.1 Organisation and dataset

In order to demonstrate the practical applicability and evaluate the effectiveness of our approach, we used the case of a major German insurer. The insurer manages an online customer service platform which contains a large number of customer problems and their corresponding solutions provided by service employees. Although many cases are available and searched in the platform by several users, a large number of new customer problems appear on a daily basis. Thus, assisted systems for employees are needed to retrieve helpful knowledge from already solved customer problems. We apply our novel hybrid three-step-approach to adequately assist the employees with a fast, correct and at the same time consistent and high quality knowledge retrieval. As the online customer service platform is hierarchically organised by topics and subtopics it shows the typical structure of these platforms. Moreover, a customer problem from the platform comprises structured and textual information which is a further characteristic property of online customer service platforms. For these reasons, the corresponding platform provides an appropriate setting to apply our novel hybrid approach. The online customer service platform of the German insurer contains knowledge in terms of 9.625 cases belonging to 26 topics including 145 subtopics. A case comprises domain categories, a customer question and a solution as depicted in Figure 1. Furthermore, we received a list of 651 insurance-specific terms (domain-specific information) and their synonyms (Information Entities) from the German insurer to initialize nodes and edges in the Case Retrieval Network. In addition, we access a Thesaurus (Naber, 2004; Naber, 2005) with over 150,000 words to regard further semantic relationships between general terms.

4.2 Demonstration of our approach at the insurer

In the following, as an essential part of the Design Science research process (cf., e.g. Peffers et al., 2007), we demonstrate the applicability of our approach. To do so, we focus on a case base consisting of 821 pre-processed cases belonging to the topics “Disability Insurance” and “Private Pension Insurance” and all corresponding subtopics. The purpose of pre-processing is to extract relevant information and to reduce the amount of terms from the customer questions to the minimum of relevant words (Abrahams et al., 2012; Li and Wu, 2010; Park and An, 2010; Storn and Price, 1997; Tralli et al., 2001). Furthermore, our case base contains hierarchically indexed cases based on the domain categories. In addition, the terms of the pre-processed customer questions are represented by their respective set of synonyms, since terms with the same semantic meaning can be clustered. Terms without a corresponding synonym are provided as the result of further pre-processing (Abrahams et al., 2012; Li and Wu, 2010; Park and

An, 2010; Sebastiani, 2002; Tralli et al., 2001). Thus, we apply pre-processing insofar as terms of the customer question had been cleared from stopwords, transformed to lower case and reduced to their word stems. To adequately determine the overall similarity between the cases in the case base and the customer problem, incoming customer problems were pre-processed in the same manner. Based on the pre-processed and hierarchically organised cases in the case base we applied our three-step-approach.

Following the *first step* of our approach, we used the hierarchically organised domain categories of the customer problem (structured information) for the Typification Tree retrieval. Following our example in Figure 1, the domain categories “Disability Insurance” and “Payment” represent a topic and a subtopic. Those categories are used to shrink the case base to cases indexed by the same domain categories. By first regarding the topic “Disability Insurance”, the initial amount of 9.625 cases in the online customer service platform is reduced to 584 cases containing the same topic. Further, the subtopic “Payment” again shrank the amount of relevant cases from 584 to only 78 cases, representing the result set $T(P_C)$. This reduction demonstrates the impact of the structural knowledge retrieval as the overall similarity determination in the next steps is dependent on the information contained in the customer questions of the set $T(P_C)$. If a customer adds additional information by specifying a second subtopic “Deadline”, the result set would be extended by 38 cases belonging to the subtree of the additional subtopic.

For the *second step*, a Case Retrieval Network is required to determine the domain-specific weighting. We initialised the Case Retrieval Network using the provided list of 651 insurance-specific terms and their synonyms as Information Entities and all cases from the case base. To do so, we connected each Information Entity to all other Information Entities through similarity arcs and each case to its Information Entities through relevance arcs. The value of semantic similarity $sim_S(e_i, e_j)$ representing the similarity arcs is set to one, if two Information Entities are synonyms and zero otherwise. Additionally, the value of the relevance arc is set to one, if the case contains a considered Information Entity and zero otherwise. In order to automatically identify Information Entities within the customer problem, the list of insurance-specific terms and their synonyms is compared against the terms of the pre-processed customer problem. To perform the second step, the network similarity $sim_N(P_Q, B_Q)$ for the customer problem and each case in the result set $T(P_C)$ has to be calculated as basis for the domain-specific weighting. Consider, for instance, a customer problem containing the Information Entities “accident” and “disabled” and a case containing their synonyms “crash” and “invalid”. Comparing each Information Entity from the customer problem with each Information Entity in the case, we calculate the maximum semantic similarities (e.g., $sim_S("disabled", "invalid") = 1$). Therefore, the network similarity yields $sim_N(P_Q, B_Q) = \frac{\max(1,0)+\max(0,1)}{2} = 1$. To keep the Case Retrieval Network current, a case comprising the customer problem with its corresponding solution is added and connected to the network within the Retain phase of the CBR cycle. By doing so, relevance arcs, obtaining the value one, are added to the network connecting the identified Information Entities with the case.

In order to perform the *third step,* a term weighting function $VSM_{T(P_C)}(\,)$ and a similarity measure $sim_{\text{VSM}}(\,)$ are required. While the parameters can also be set or optimised using machine learning techniques (Cortes and Vapnik, 1995; Cover and Hart, 1967; John and Langley, 1995; Park et al., 1991; Quinlan, 1986), we apply a fixed parametrisation as a starting point. Thereby, we decided to use $tf.idf$ for the function $VSM_{T(P_C)}(\,)$, since it is the most widely used statistical term weighting method for text analysis (Hua et al., 2009; Salton and Buckley, 1988; Sebastiani, 2002). Subsequently, the customer questions of the customer problem P_Q and each customer question B_Q from the shrunken case base $T(P_C)$ are represented as the vectors v_p and v_b in the VSM by applying $tf.idf$ (cf. Equation (5)). Hence, we obtain vector entries v_p^i and v_b^i by weighting the frequency of term occurrences in the customer question of the customer problem $tf(t_i^{syn}, P_Q)$ or the case $tf(t_i^{syn}, B_Q)$ with the inverse frequency of term occurrences over all cases of the shrunken case base $idf\left(t_i^{syn}, T(P_C)\right)$.

$$v_p^i = tf(t_i^{syn}, P_Q) * idf\left(t_i^{syn}, T(P_C)\right), v_b^i = tf(t_i^{syn}, B_Q) * idf\left(t_i^{syn}, T(P_C)\right), \forall\, i = 1, \dots, n \tag{5}$$

In addition to the statistical $tf.idf$ weighting, the entries of the vectors v_p and v_b are transformed based on the domain-specific weighting from the second step by increasing vector entries of terms which correspond to the identified Information Entities (cf. Equation (3)). To determine the overall similarity between the customer problem and each case, we used the cosine similarity measure $sim_{\cos(\theta)}\big(w_p(P_Q, B_Q), w_b(P_Q, B_Q)\big)$ which is commonly used for similarity determination in vector space representations (Salton and McGill, 1983). The cosine measure bases on the angle between the modified vectors w_p and w_b with possible values in the interval $[0; 1]$ (cf. Equation (6)).

$$SIM(P,B) = sim_{\cos(\theta)}\big(w_p(P_Q, B_Q), w_b(P_Q, B_Q)\big), = \frac{\sum_{i=1}^{n} w_p^i(P_Q, B_Q) * w_b^i(P_Q, B_Q)}{\sqrt{\sum_{i=1}^{n}\big(w_p^i(P_Q, B_Q)\big)^2} * \sqrt{\sum_{i=1}^{n}\big(w_b^i(P_Q, B_Q)\big)^2}} \forall\, B \in T(P_C) \qquad (6)$$

In sum, to attain a ranking of all cases in the shrunken case base $T(P_C)$, we calculated the overall similarity for each pair of customer problem and case. Hence, an employee of an online customer service platform is provided with helpful knowledge contained in the solutions of the most similar cases.

4.3 Evaluation

In order to evaluate the quality of the results of our approach, we compared the cases retrieved by our approach to cases assigned as helpful from domain experts of the insurer. More precisely, the German insurer provided us for the purpose of evaluation with 87 additional customer problems, referred to as test set, containing requests towards knowledge within the case base with respect to the topic “Disability Insurance”. Subsequently, domain experts manually linked these 87 customer problems to cases in the case base, which were considered helpful. As a result, each customer problem of the test set was assigned at least one helpful case from the case base. On this basis, we were able to rigorously evaluate knowledge retrieval approaches in our setting. Knowledge retrieval approaches aim to provide the employees with the most helpful cases to solve a new customer problem. Since employees cannot search huge sets of potentially helpful cases, knowledge retrieval approaches provide a fixed maximum number of *k* retrieved cases. In order to evaluate the number of correctly retrieved cases, related research areas commonly refer to the well-established measures precision and recall (Forman, 2003; Sokolova et al., 2006). However, research in CBR emphasizes the importance of retrieving at least one relevant case rather than measuring the proportion of retrieved relevant cases regarding all relevant cases (recall) or all retrieved cases (precision) (Burke et al. 1998; Lenz et al. 1998). Thus, in line with CBR literature, we refer the knowledge retrieval for a specific customer problem as successful if at least one helpful case is contained within the *k* retrieved cases of the respective approach. From this point of view, we also increase comparability to competing artifacts from CBR literature since a successful knowledge retrieval in our sense constitutes the adapted definition of recall from Lenz et al. (1998). To quantify the quality of the results, we refer to the proportion of customer problems for which knowledge retrieval was successful with respect to the total number of customer problems in the test set. In the following, we refer to this measure as proportion of *successful retrievals*. To analyse the quality of the results for different maximum numbers of retrieved cases *k*, we started with $k = 0$ and incrementally increased this number by one.

In a first step, we compared the quality of the results of our approach with the quality of the results of other basic state-of-the-art approaches and random sampling without replacement as baseline approach (Horvitz and Thompson, 1952). Since the well-known approaches CBR-Answers (Lenz and Burkhard, 1996; Lenz and Burkhard, 1997; Lenz et al., 1998; Lenz et al., 1999) and FAQ Finder (Burke et al., 1997; Hammond et al., 1995) are still state-of-the-art from a methodical point of view, we refer to these approaches as baselines for comparison. By doing so, we are in line with recent research (Balakrishnan and Zhang, 2014; Shekhar et al., 2014; Sizov et al., 2015). In contrast, pure structured approaches like Typification Tree retrieval (Heras et al., 2009) do not aim at analysing textual information. Furthermore, the hierarchical structure defined by the domain categories in online customer service platforms generally does not provide sufficient depth to select an appropriate number of *k* cases. Therefore, we were not able to compare our approach with pure structured approaches. Figure 3 depicts the results of the analysis of our hybrid knowledge retrieval approach in comparison with the mentioned state-of-the-art approaches and random sampling without replacement. Thereby, the proportion of successful retrievals is plotted for each maximum number of retrieved cases k. Regarding the comparison with competing artifacts, shown on the left hand side in Figure 3, our hybrid approach performed 45.98% (65.52%) of the retrievals successfully, when retrieving only one (two) case(s). In contrast, the FAQ Finder reached a value of 27.59% (40.23%) successful retrievals. With 18.39% (25,29%) the CBR-Answers approach had the second lowest value for successful retrieval. At last, the random sampling without replacement baseline performed worst by only retrieving 3.79% (7.47%) of cases with success. Since a set of five cases should be fairly manageable on first sight by service professionals, there is an about 80% chance of finding helpful cases by our hybrid approach for this maximum number of retrieved cases.

In a second step, to gain more detailed insights with respect to our three-step-approach combining Typification Tree, Case Retrieval Network and VSM we compared the quality of its results with the quality of the results of the respective twofold combinations Typification Tree and Case Retrieval Network, Typification Tree and VSM as well as Case Retrieval Network and VSM. In comparison with the twofold method combinations on the right hand side of Figure 3, our three-step-approach outperformed the others for low values of k, which are the most important to provide service professionals quickly with helpful cases. Regarding the combination of Typification Tree and Case Retrieval Network for $k > 8$, there are no major differences to the proposed approach. The combination of Case Retrieval Network and VSM provided the worst results. Nevertheless, the results foster our argumentation to combine structured and textual information, since the twofold method combinations including the Typification Tree outperform combinations not incorporating structured information. Of course, the instantiation of our approach in a practical context goes along with respective costs. As the instantiation of our approach has only to take place once for a given context of application to benefit from its advantages, these costs are primarily initial effort. Major part of this initial effort may be due to the domain-specific weighting by the network component. However, these costs can be kept low if Information Entities are extracted in an automated way and/or only synonyms of terms are considered regarding the semantic similarity in the network. Summing up, the results of our comparisons reveal that our hybrid approach provides higher quality than the competing artifacts and the twofold method combinations.

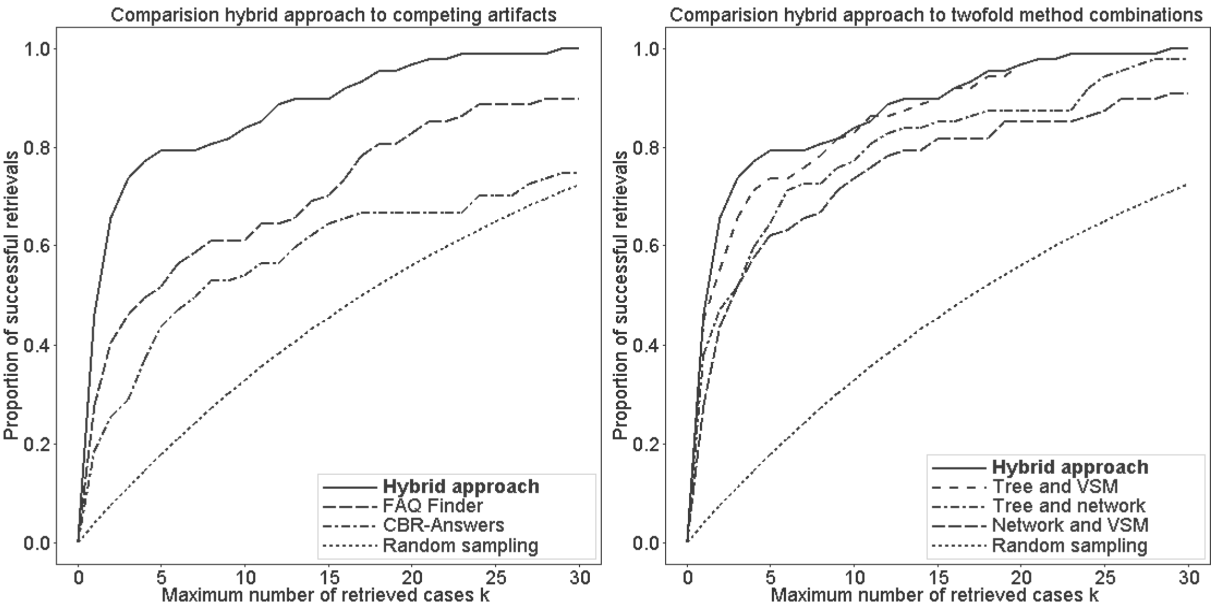

Figure 3: Evaluation of our approach in comparison with competing artifacts (left) and the twofold method combinations (right).

5 Conclusion, Limitations and Future Research

Nowadays, organisations face the challenge of a high quality and consistent knowledge retrieval for employees in online customer service platforms to provide solutions for customer problems. Since customer problems comprise structured as well as textual information, CBR provides promising approaches for an automated knowledge retrieval (Barão et al., 2017; Martin et al., 2017; Lee et al., 2015) dealing with the two types of information. Nevertheless, until now literature does not provide sufficient approaches combining structural (Burke et al., 2000; Heras et al., 2009; Ricci and Senter, 1998) and textual (Burke et al., 1997; El-Sappagh and Elmogy, 2015; Lenz et al., 1998; Shekhar et al., 2014) CBR. Therefore, we propose a new hybrid knowledge retrieval approach combining structured and textual information in an integrated perspective to adequately assist employees in online customer service platforms.

Our hybrid approach comprises three steps. First, we make use of the hierarchical structured information in online customer service platforms to shrink the case base to a subset of relevant cases regarding the domain categories of the customer problem. Second, a domain-specific weighting between the customer problem and each case of the shrunken case base is determined to reward cases containing the same or nearly the same domain-specific information as the customer problem. Third, we determine the overall similarity of the customer problem to the subset of relevant cases by considering general information contained in the customer problem along with the domain-specific weighting. As a result, the most similar cases regarding the customer problem are retrieved in order to provide helpful knowledge. We demonstrated and evaluated our approach using the case of a German insurer. The results of the evaluation reveal that our hybrid knowledge retrieval approach provides higher quality compared to other basic state-of-the-art approaches and random sampling without replacement. Moreover, the integration of all three steps outperforms all twofold combinations.

From a theoretical perspective, our work contributes to literature by taking an integrated perspective in structural CBR as well as textual CBR. Since literature does not provide a sufficient combination of structural and textual CBR, we address this gap. The implications of our hybrid approach in the field of customer service in online customer service platforms are two folded. First, we outline that employees, dealing with an excessive amount of customer problems in online customer service platforms, can be supported in an appropriate way. Indeed, our approach provides helpful solutions for new customer problems in an automated way based on similar already solved customer problems. Second, consistency is ensured since all employees are provided with the same knowledge, usually distributed within an organisation. We are confident, that our proposed approach will help organisations to significantly improve their customer service and reduce overall costs. Besides these benefits, our work is also subject to some limitations. First, we only considered one online customer service platform from the insurance domain, for which we applied and evaluated our approach. However, we chose the insurance domain as it is a very complex and challenging field. Nevertheless, we encourage further research and respective applications in other domains. Second, while in a first step, we focused on the Retrieve phase of the CBR cycle, it seems promising to also investigate the other phases of the CBR cycle. Since our hybrid approach aims at retrieving the most helpful cases for service employees, we recommend future research on approaches to adequately combine helpful cases providing an advanced knowledge representation. Going a step further, the question arises how a knowledge retrieval approach could be extended and developed further to automatically solve customer problems in online customer service platforms without the need of any employees. As a promising starting point, it seems reasonable to apply machine learning techniques reducing the initial costs of the first step (assigning domain categories) and the second step (extracting Information Entities). Summing up, we believe that our study is an important step in terms of the combination of structural and textual knowledge retrieval in the area of online customer service platforms. We hope our work will stimulate further research in this exciting field.

References

Aamodt, A. (1991). “A knowledge-intensive, integrated approach to problem solving and sustained learning”. PhD thesis. University of Trondheim.

Aamodt, A. and E. Plaza (1994). “Case-based reasoning. Foundational issues, methodological variations, and system approaches” *AI communications* 7 (1), 39–59.

Abrahams, A. S., J. Jiao, G. A. Wang and W. Fan (2012). “Vehicle Defect Discovery From Social Media” *Decision Support Systems* 54 (1), 87–97.

Acorn, T. L. and S. H. Walden (1992). "SMART. Support management automated reasoning technology for Compaq customer service". In: *Proceedings of the 4th conference on Innovative applications of artificial intelligence*: AAAI Press, pp. 3–18.
Alavi, M. and D. E. Leidner (1999). "Knowledge management systems. Issues, challenges, and benefits" *Communications of the AIS* 1 (7).
Alavi, M. and D. E. Leidner (2001). "Knowledge management and knowledge management systems. Conceptual foundations and research issues" *MIS quarterly* 25 (1), 107–136.
Aleven, V. (1997). *Teaching Case-Based Argumentation Through a Model and Examples*. PhD thesis. University of Pittsburgh.
Aleven, V. and K. D. Ashley (1997). "Teaching case-based argumentation through a model and examples empirical evaluation of an intelligent learning environment". In B. Du Boulay and R. Mizoguchi (eds.) *Artificial Intelligence in Education. Knowledge and Media in Learning Systems*, pp. 87–94: IOS press.
Ashley, K. D. (1991). "Reasoning with cases and hypotheticals in HYPO" *International Journal of Man-Machine Studies* 34 (6), 753–796.
Balakrishnan, V., K. Ahmadi and S. D. Ravana (2016). "Improving retrieval relevance using users' explicit feedback" *Aslib Journal of Information Management* 68 (1), 76–98.
Balakrishnan, V. and X. Zhang (2014). "Implicit user behaviours to improve post-retrieval document relevancy" *Computers in Human Behavior* 33, 104–112.
Barão, A., J. B. de Vasconcelos, Á. Rocha and R. Pereira (2017). "A knowledge management approach to capture organizational learning networks" *International Journal of Information Management* 37 (6), 735–740.
Behbahani, M., A. Saghaee and R. Noorossana (2012). "A case-based reasoning system development for statistical process control. Case representation and retrieval" *Computers & Industrial Engineering* 63 (4), 1107–1117.
Bergmann, R. (2002). *Experience Management. Foundations, Development Methodology, and Internet-Based Applications:* Springer.
Bergmann, R. and A. Stahl (1998). "Similarity measures for object-oriented case representations" *Advances in Case-Based Reasoning*, 25–36.
Bergmann, R. and W. Wilke (1996). "On the role of abstraction in case-based reasoning" *Advances in Case-Based Reasoning*, 28–43.
Brüninghaus, S. and K. Ashley (2005). "Reasoning with Textual Cases". In: *Proceedings of the 4th International Conference on Case-Based Reasoning*, pp. 137–151.
Brüninghaus, S. and K. D. Ashley (1997). "Using machine learning for assigning indices to textual cases". In: *Proceedings of the International Conference on Case-Based Reasoning*, pp. 303–314.
Brüninghaus, S. and K. D. Ashley (1999). "Bootstrapping Case Base Development with Annotated Case Summaries⋆". In: *Proceedings of the International Conference on Case-Based Reasoning*, pp. 59–73.
Burke, E. K., B. MacCarthy, S. Petrovic and R. Qu (2000). "Structured cases in case-based reasoning—re-using and adapting cases for time-tabling problems" *Knowledge-based systems* 13 (2), 159–165.
Burke, R. D., K. J. Hammond, V. Kulyukin, S. L. Lytinen, N. Tomuro and S. Schoenberg (1997). "Question answering from frequently asked question files. Experiences with the FAQ FINDER system" *AI magazine* 18 (2), 57–66.
Campos, R., G. Dias, A. M. Jorge and A. Jatowt (2015). "Survey of temporal information retrieval and related applications" *ACM Computing Surveys (CSUR)* 47 (2), 15.
Carpineto, C. and G. Romano (2012). "A survey of automatic query expansion in information retrieval" *ACM Computing Surveys (CSUR)* 44 (1), 1.
Chen, J.-g. and Y.-z. Chen (2011). "Research on operation standard expert system based on ontology and CBR". In: *Proceedings of the 18th International Conference on Industrial Engineering and Engineering Management (IE&EM)*, pp. 1932–1936.
Cortes, C. and V. Vapnik (1995). "Support-Vector Networks" *Machine learning* 20 (3), 273–297.

Cover, T. and P. Hart (1967). “Nearest Neighbor Pattern Classification” *IEEE transactions on information theory* 13 (1), 21–27.

Cunningham, C., R. Weber, J. M. Proctor, C. Fowler and M. Murphy (2004). “Investigating graphs in textual case-based reasoning” *Advances in Case-Based Reasoning*, 573–586.

Daniels, J. J. and E. L. Rissland (1997). “Integrating IR and CBR to locate relevant texts and passages”. In: *Proceedings of the 8th International Workshop on Database and Expert Systems Applications*, pp. 206–212.

de Mantaras, R. L., D. McSherry, D. Bridge, D. Leake, B. Smyth, S. Craw, B. Faltings, M. L. Maher, M. T. Cox, K. Forbus, M. Keane, A. Aamodt and I. Watson (2005). “Retrieval, reuse, revision and retention in case-based reasoning” *The Knowledge Engineering Review* 20 (3), 215–240.

Decker, B. (2016). *Life In A Fast Paced Society*. *Have a slow down moment within our fast lives.* URL: https://www.theodysseyonline.com/life-in-fast-paced-society (visited on 11/27/2017).

El-Sappagh, S. H. and M. Elmogy (2015). “Case Based Reasoning. Case Representation Methodologies” *International Journal of Advanced Computer Science and Applications,(IJACSA)* 6 (11), 192–208.

eMarketer (2016). *Customer Service Channels That Frustrate Consumers*. *Consumers will abandon a retailer because of poor customer service*. URL: https://www.emarketer.com/Article/Customer-Service-Channels-That-Frustrate-Consumers/1013637 (visited on 11/27/2017).

Forman, G. (2003). “An Extensive Empirical Study of Feature Selection Metrics for Text Classification” *Journal of machine learning research* 3 (Mar), 1289–1305.

Forrester (2016). *Your Customers Don’t Want To Call You For Support*. URL: https://go.forrester.com/blogs/16-03-03-your_customers_dont_want_to_call_you_for_support/ (visited on 11/27/2017).

Göker, M. H. and T. Roth-Berghofer (1999). “The development and utilization of the case-based helpdesk support system HOMER” *Engineering Applications of Artificial Intelligence* 12 (6), 665–680.

Grant, R. M. (1996). “Prospering in Dynamically-Competitive Environments. Organizational Capability as Knowledge Integration” *Organization Science* 7 (4), 375–387.

Guo, Y., J. Hu and Y. Peng (2012). “A CBR system for injection mould design based on ontology. A case study” *Computer-Aided Design* 44 (6), 496–508.

Hammond, K., R. Burke, C. Martin and S. Lytinen (1995). “FAQ finder. A case-based approach to knowledge navigation”. In: *Proceedings of the 11th Conference on Artificial Intelligence for Applications*, pp. 80–86.

Heidelberger, C. A. and S. Sarnikar (2008). “Knowledge Flow in Online Communities: A Study of the Relationship between Knowledge Complexity, Online Collaboration, and Knowledge Flow”. In: *MWAIS 2008 Proceedings*.

Heras, S., J. Á. García-Pardo, R. Ramos-Garijo, A. Palomares, V. Botti, M. Rebollo and V. Julián (2009). “Multi-domain case-based module for customer support” *Expert Systems with Applications* 36 (3), 6866–6873.

Höffner, K., S. Walter, E. Marx, R. Usbeck, J. Lehmann and A.-C. Ngonga Ngomo (2017). “Survey on challenges of question answering in the semantic web” *Semantic Web* 8 (6), 895–920.

Horvitz, D. G. and D. J. Thompson (1952). “A generalization of sampling without replacement from a finite universe” *Journal of the American Statistical Association* 47 (260), 663–685.

Hua, J., W. D. Tembe and E. R. Dougherty (2009). “Performance of Feature-Selection Methods in the Classification of High-Dimension Data” *Pattern Recognition* 42 (3), 409–424.

John, G. H. and P. Langley (1995). “Estimating Continuous Distributions in Bayesian Classifiers”. In: *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*: Morgan Kaufmann Publishers Inc., pp. 338–345.

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

Kriegsman, M. and R. Barletta (1993). “Building a case-based help desk application” *IEEE Expert* 8 (6), 18–26.

Kunze, M. and A. Hübner (1998). “CBR on semi-structured documents. The experience book and the FAllQ project”. In: *Proceedings of 6th German Workshop on CBR*.

Lee, C.-H., Y.-H. Wang and A. J. C. Trappey (2015). “Ontology-based reasoning for the intelligent handling of customer complaints” *Computers & Industrial Engineering* 84, 144–155.
Lenz, M., B. Bartsch-Spörl, H.-D. Burkhard and S. Wess (eds.) (1998). *Case-based reasoning technology: From Foundations to Applications:* Springer.
Lenz, M. and H.-D. Burkhard (1996). “Case retrieval nets. Basic ideas and extensions” *Ki-96: Advances in artificial intelligence*, 227–239.
Lenz, M. and H.-D. Burkhard (1997). “CBR for document retrieval. The FA ll Q project” *Case-Based Reasoning Research and Development*, 84–93.
Lenz, M., K.-H. Busch, A. Hübner and S. Wess (1999). “The simatic knowledge manager”. In: *Exploring Synergies of Knowledge Management and Case-Based Reasoning. Papers from the AAAI Workshop, Technical Report WS-99-10*.
Li, N. and D. D. Wu (2010). “Using Text Mining and Sentiment Analysis for Online Forums Hotspot Detection and Forecast” *Decision Support Systems* 48 (2), 354–368.
Liao, T. W., Z. Zhang and C. R. Mount (1998). “Similarity measures for retrieval in case-based reasoning systems” *Applied Artificial Intelligence* 12 (4), 267–288.
Martin, A., S. Emmenegger, K. Hinkelmann and B. Thönssen (2017). “A viewpoint-based case-based reasoning approach utilising an enterprise architecture ontology for experience management” *Enterprise Information Systems* 11 (4), 551–575.
Martin, F. J. and E. Plaza (2004). “Ceaseless case-based reasoning”. In: *Proceedings of the European Conference on Case-Based Reasoning*, pp. 287–301.
McKinsey & Company (2016). *Customer experience. Creating value through transforming customer journeys*. URL: https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/customer-experience-creating-value-through-transforming-customer-journeys (visited on 11/27/2017).
Moreo, A., M. Navarro, J. L. Castro and J. M. Zurita (2012). “A high-performance FAQ retrieval method using minimal differentiator expressions” *Knowledge-based systems* 36 (Supplement C), 9–20.
Naber, D. (2004). “Openthesaurus. Building a thesaurus with a web community” *Retrieved January* 3, 2005.
Naber, D. (2005). “OpenThesaurus. Ein offenes deutsches Wortnetz” *Sprachtechnologie, mobile Kommunikation und linguistische Ressourcen: Beiträge zur GLDV-Tagung, Bonn, Germany*, 422–433.
Noble, B. (2016). *The $62 billion customer service scared away [INFOGRAPHIC]*. URL: https://www.newvoicemedia.com/blog/the-62-billion-customer-service-scared-away-infographic (visited on 11/27/2017).
Oracle Corporation (2011). “2011 Customer Experience Impact Report. Getting to the Heart of the Consumer and Brand Relationship”.
Park, D. C., M. A. El-Sharkawi, R. J. Marks, Le Atlas and M. J. Damborg (1991). “Electric Load Forecasting Using an Artificial Neural Network” *IEEE transactions on Power Systems* 6 (2), 442–449.
Park, S. and D. U. An (2010). “Automatic E-Mail Classification Using Dynamic Category Hierarchy and Semantic Features” *IETE Technical Review* 27 (6), 478–492.
Peffers, K., T. Tuunanen, M. A. Rothenberger and S. Chatterjee (2007). “A Design Science Research Methodology for Information Systems Research” *Journal of Management Information Systems* 24 (3), 45–77.
Plaza, E. (1995). “Cases as terms. A feature term approach to the structured representation of cases”. In: *Proceedings of the 1st International Conference on Case-Based Reasoning*, pp. 265–276.
Prince, W. W. (2017). *Reference for Business*. *KNOWLEDGE WORKERS*. URL: http://www.referenceforbusiness.com/management/Int-Loc/Knowledge-Workers.html (visited on 11/27/2017).
Quinlan, J. R. (1986). “Induction of Decision Trees” *Machine learning* 1 (1), 81–106.
Recio, J., B. Diaz-Agudo, M. Gomez-Martin and N. Wiratunga (2005). “Extending jCOLIBRI for textual CBR”. In: *Proceedings of the 6th International Conference on Case-Based Reasoning*, pp. 421–435.

Ricci, F. and L. Senter (1998). “Structured cases, trees and efficient retrieval”. In: *Proceedings of the 4th European Workshop on Case-based Reasoning*, pp. 88–99.

Rubenstein-Montano, B., J. Liebowitz, J. Buchwalter, D. McCaw, B. Newman, K. Rebeck and T. K. M. M. Team (2001). “A systems thinking framework for knowledge management” *Decision Support Systems* 31 (1), 5–16.

Salton, G. and C. Buckley (1988). “Term-weighting approaches in automatic text retrieval” *Information processing & management* 24 (5), 513–523.

Salton, G. and M. J. McGill (1983). *Introduction to Modern Information Retrieval.* 1st Edition. New York: McGraw-Hill, Inc.

Salton, G., A. Wong and C.-S. Yang (1975). “A vector space model for automatic indexing” *Communications of the ACM* 18 (11), 613–620.

Sebastiani, F. (2002). “Machine Learning in Automated Text Categorization” *ACM Computing Surveys (CSUR)* 34 (1), 1–47.

Shekhar, S., S. Chakraborti and D. Khemani (2014). “Linking cases up. An extension to the case retrieval network”. In: *International Conference on Case-Based Reasoning*, pp. 450–464.

Shimazu, H. (1998). “A textual case-based reasoning system using XML on the world-wide web”. In: *Proceedings of the 4th European Workshop on Case Based Reasoning*, pp. 274–285.

Simoudis, E. (1992). “Using case-based retrieval for customer technical support” *IEEE Expert* 7 (5), 7–12.

Sizov, G., P. Öztürk and A. Aamodt (2015). “Evidence-driven retrieval in textual CBR. Bridging the gap between retrieval and reuse”. In: *International Conference on Case-Based Reasoning*, pp. 351–365.

Sokolova, M., N. Japkowicz and S. Szpakowicz (2006). “Beyond Accuracy, F-score and ROC: A Family of Discriminant Measures for Performance Evaluation”. In: *Australasian Joint Conference on Artificial Intelligence*. Heidelberg: Springer Berlin, pp. 1015–1021.

Sprout Social, I. (2016). *The Q2 2016 Sprout Social Index*. URL: https://sproutsocial.com/insights/data/q2-2016/ (visited on 11/27/2017).

Stahl, A. (2005). “Learning similarity measures. A formal view based on a generalized CBR model”. In: *International Conference on Case-Based Reasoning*, pp. 507–521.

Stahl, A. and T. Gabel (2006). “Optimizing similarity assessment in case-based reasoning”. In: *Proceedings of the international conference on artificial intelligence*, p. 1667.

Statista (2017). *Number of internet users worldwide from 2005 to 2017 (in millions)*. URL: https://www.statista.com/statistics/273018/number-of-internet-users-worldwide/ (visited on 11/27/2017).

Storn, R. and K. Price (1997). “Differential Evolution – A Simple and Efficient Heuristic for Global Optimization Over Continuous spaces” *Journal of global optimization* 11 (4), 341–359.

Tiwana, A. (2000). *The knowledge management toolkit. Practical techniques for building a knowledge management system:* Prentice Hall PTR.

Tralli, V., A. Vaccari, R. Verdone and O. Andrisano (2001). “Adaptive Time and Frequency Resource Assignment with COFDM for LMDS Systems” *IEEE Transactions on Communications* 49 (2), 235–238.

Turtle, H. and W. B. Croft (1989). “Inference networks for document retrieval”. In: *Proceedings of the 13th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 1–24.

Watson, I. (1999). “Case-based reasoning is a methodology not a technology” *Knowledge-based systems* 12 (5), 303–308.

Watson, I. and S. Perera (1998). “A hierarchical case representation using context guided retrieval” *Knowledge-based systems* 11 (5-6), 285–292.

Weber, R. O., K. D. Ashley and S. Brüninghaus (2005). “Textual case-based reasoning” *The Knowledge Engineering Review* 20 (3), 255–260.

Weis, K.-H. (2015). “A Case Based Reasoning Approach for Answer Reranking in Question Answering”. In: *Proceedings Informatik*.

Wilson, D. C. and S. Bradshaw (1999). “CBR textuality”. In: *Proceedings of the 4th UK Case-Based Reasoning Workshop*, pp. 67–80.

Yan, A., L. Qian and C. Zhang (2014). “Memory and forgetting: An improved dynamic maintenance method for case-based reasoning” *Information Sciences* 287, 50–60.

Yokoyama, T. (1990). “An object-oriented and constraint-based knowledge representation system for design object modeling”. In: *Proceedings of the 6th Conference on Artificial Intelligence Applications, 1990*, pp. 146–152.

Zendesk, I. (2017). *Why companies should invest in the customer experience*. URL: https://www.zendesk.com/resources/why-companies-should-invest-in-the-customer-experience/ (visited on 11/27/2017).

Zhang, N., C. Wang and Y. Xu (2011). “Privacy in Online Social Networks”. In: *ICIS 2011 Proceedings*.