Case-Based Reasoning: Potential Benefits and Limitations for Documenting of Stories in Organizations

Keywords

Case-Based Reasoning, Storytelling, Organizations, Experience Management

Abstract

This article investigates the combination of two knowledge management approaches, Storytelling, especially Triad talks, and Case-Based Reasoning. First, we shortly introduce the approaches and then describe in more detail how to combine these approaches. In addition, we give an overview of the potential benefits and limitations of the combination approach. Finally we summarize the article and give an outlook to future work.

Practical Relevance

By combining different knowledge management approaches, the knowledge conservation and transfer within organizations can be designed in a more efficient way. The strengths of different approaches can be combined and the shortcomings can be minimized. Triad talks as well as Case-Based Reasoning have proven their practical usability, and therefore a combination of these approaches could be an advantage for organizations. This article will illustrate the possibilities, potential benefits, and limitations and show the practical usability.

1 Introduction

Experience Management in organizations is an important task. Several different approaches exist to identify, extract and store the experience of employees in an organization. Some approaches are more analogous like Interviews, Knowledge Maps, Storytelling, and Workshops, while others are more digital like Wikis, Document storages, and several technologies from artificial intelligence.

In this article we investigate the combination of an analogous approach, Storytelling and the special form Triad talks, with a digital approach, Case-Based Reasoning and the potential benefits and limits of this combination.

Storytelling is an experience management approach and is used to create so-called experience stories. An experience story contains information about important events from the past of an organization. The story comprises subjective experience from the interviewed employees as well as additional reflective and context-sensitive comments from the interviewer. The goal of the storytelling approach is to preserve and distribute the experience and the implicit knowledge from experts to make the gained knowledge accessible for the organization. The narrative and analogous character of the storytelling is able to trigger change and learning processes. (ReinmannRothmeier et al. 2000; Stary et al 2013)

Triad talks are a special form of the Storytelling and the approach was developed by Michael Dick in 2006. Triad talks are interviews between three participants. The talks are limited in respect to the location, the time, and the discussed topic. The role of the participants is defined at the beginning of the interview. The first participant is the expert and experience medium. He is the narrator during the interview. The second participant is a novice to the topic of the interview and he wants to learn from the expert. The third participant is the so-called methodical listener. He is also a novice to the topic of the interview, but his role is to assure the understanding between the expert and the novice. He encourages the expert to explain the experiences in detail, asks for clarifications, and moderates the interview. During the interview, the methodical listener also writes down the questions and their answers to store the results after the interview. He is also responsible for the reuse of the results. (Dick 2006)

While Storytelling and Triad talks are analogous approaches to experience management, Case-Based Reasoning (CBR, in German: Fallbasiertes Schließen) is a more digital approach to make experience storeable and accessible through a computer. CBR is a problem solving approach from artificial intelligence and it is based on the idea that similar problems have similar solutions. Experience from the past is used and adapted to solve newly occurred problems. The experience knowledge is stored in so-called cases. The cases, which are most similar to the current problem situation, are used to find a solution. A case consists in minimum of a problem description and a corresponding solution. In addition, a case may contain information about the quality of the solution, adaptation steps or alternative solutions. All cases are stored in a so-called case base.(Kolodner 1993)

The case base is one of the four knowledge containers that are used in CBR. The other containers are the vocabulary, the similarity measures and the adaptation knowledge. The vocabulary contains all terms that are used within a CBR system, while the similarity measures are used to define how to compute the similarity between a new problem and the existing cases and to approximate the usefulness of an existing case to the new problem. The adaptation knowledge contains information how to adapt the solution of existing cases to fit to the new problem. Mostly this knowledge is stored as adaptation rules. (Richter 2003)

The process-model for CBR, the so-called 4R-cylce, was developed by Aamodt and Plaza in 1994 and contains four steps: Retrieve Reuse, Revise, and Retain. Figure 1 shows this process model.

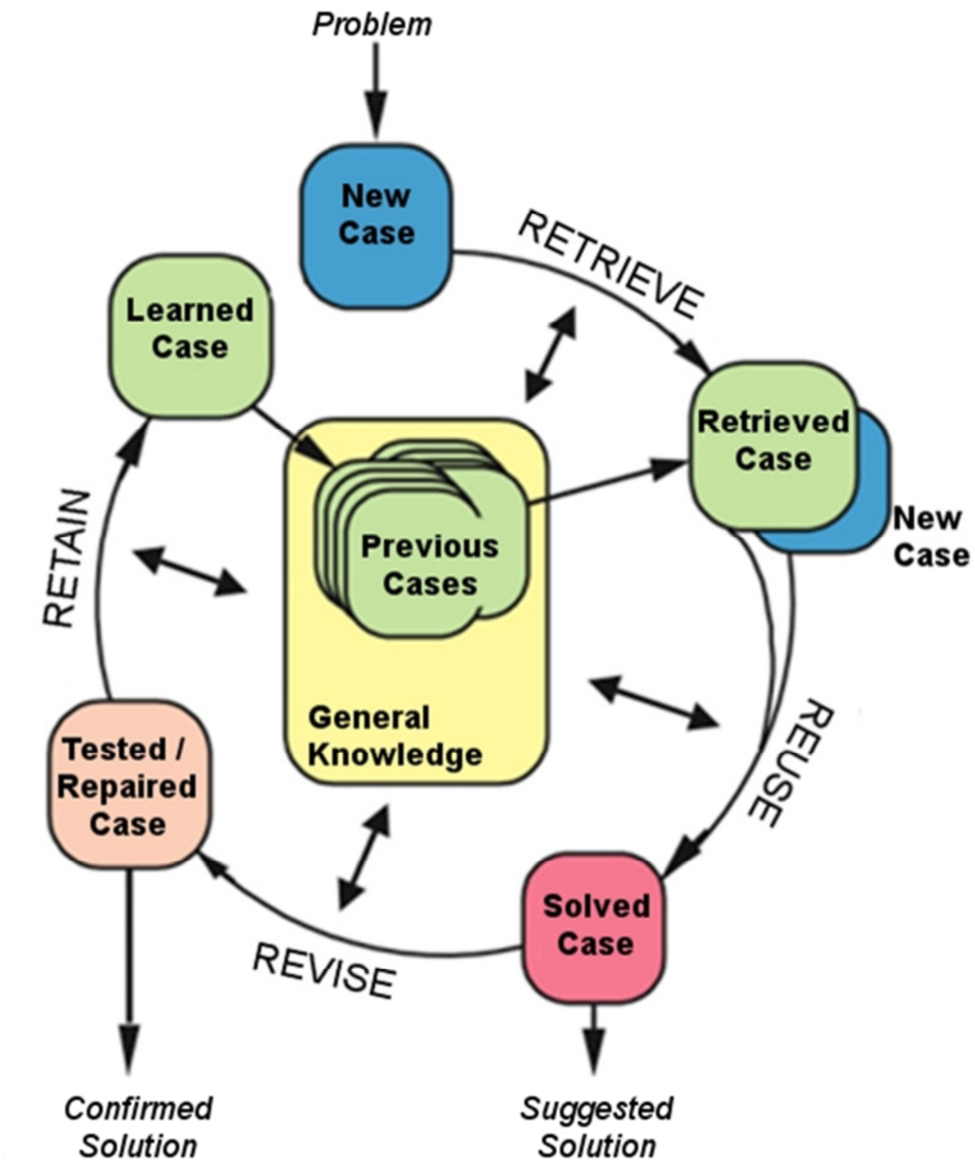

Figure 1: The CBR process model (cf. AamodtPlaza 1994)

The Retrieve step transforms the new problem into the given case structure and searches for the most similar cases in the case base. The Reuse step is responsible for adaption the solutions of the retrieved cases to the given problem. The Revise step is used to validate the adapted solutions. The validation can be done by the user of a CBR system, domain experts or with a simulation. If a solution is not valid it can be further adapted and validated again or discarded. The last step, Retain, is the learning step of a CBR system. The new problem with the adapted and validated solution can be stored in the case base to use it to solve future problems. In addition, similarity measures or adaption knowledge can be learned, too. (Aamodt Plaza 1994)

In the following, we describe a combination approach of Triad talks with Case-Base Reasoning and the potential benefits as well as the limitations of the combination approach. Finally, we give a summary and an outlook to future work.

2 Combination of Triad talks and Case-Based Reasoning

Triad talks and CBR are experience management approaches and the idea to combine them seems naturally. In both approaches the reuse of experience is the main goal. Triad talks and CBR are storing experience and try to reuse it in similar situations to solve a given problem or provide additional information for decision support. Therefore, it is necessary to determine the usefulness of a past experience. For the results of Triad talks, a human has to read the interview document and determine whether it may be useful or not. This is a time consuming process, even when dozens or hundreds of interviews have been done. In a CBR system the defined similarity measures are used to compute the similarity between cases and then the results are displayed to the user. But a CBR system is not able to act as a methodical listener and moderate an interview or write down the results.

Triad talks are very good for transforming implicit knowledge from an expert to explicit knowledge that can be used by other persons, but the reuse of interview results is challenging and resource consuming. CBR has a very good experience reuse mechanism with the defined similarity measures and can quickly handle large numbers of cases, but the initial recording of knowledge and experience has to be done by a human and can be resource intensive. Combining Triad talks with CBR takes advantage of the strengths of both approaches and can minimize the shortcomings.

The idea is to perform a Triad talk, write down the results and then transform the free text into a knowledge structure that can be used by a CBR system. The CBR system then can be used to find similar interviews or part of interviews to a newly given situation. The retrieved documents can be used to plan future Triad talks, support the methodical listener during the interview, and find connections between interviews. Figure 2 shows the steps of the combination in form of a cycle.

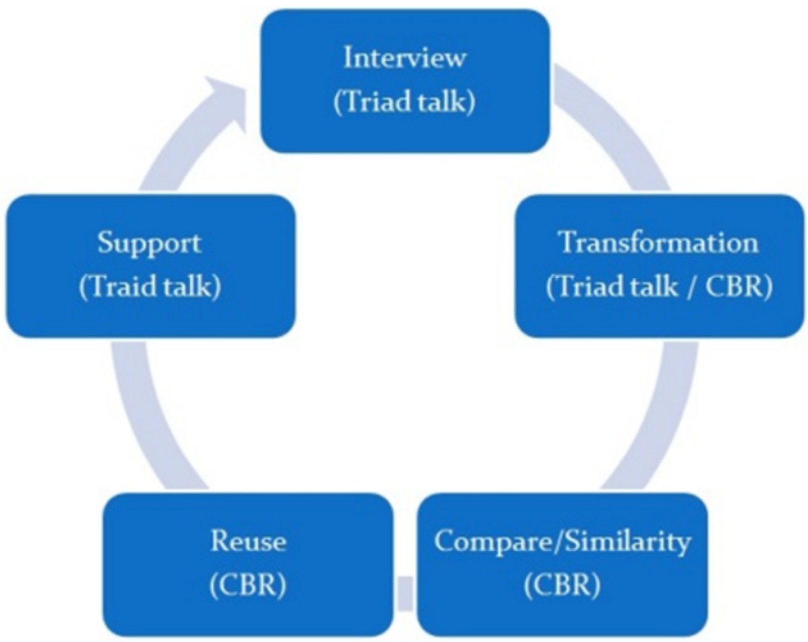

Figure 2: Steps for combining of Triad talks and CBR

The knowledge structure for the transformation step depends on the type of CBR system. There are three different types of CBR systems: textual, structural, and conversational. A textual CBR system uses text documents as cases and information retrieval techniques to compare the cases to a new situation. The similarity computation can be challenging in this type of CBR system, because the documents are unstructured and may have different sizes. (Lenz 1999) In a structural CBR system, the cases are represented as attribute-value-pairs. A case consists of several attributes with one or multiple values of different data types. The similarity is computed by comparing the attribute values of a given case and the new situation (local similarity) and combining them to an amalgamation function to compute the similarity of the whole case (global similarity). (Pfuhl 2003) A conversational CBR system represents a case as a flat list of questions. A new situation is briefly described as free text and the conversational CBR system ask questions. After every answered question, the system displays a list of relevant cases according to the answer of the question. Every answered question reduces the list of cases until the most relevant case is found. (Aha 2001)

While the textual approach seems a good starting point for combining Triad talks and CBR, because an interview document may be used as a case, the similarity computation between these documents will be very difficult. The interview documents tend to be different in terms of structure, size, and content, and therefore the information retrieval mechanism can only be implemented with high effort. In addition, an interview document may contain information about different topics or sub-topics and it would be more appropriate to split the knowledge and experience of the document into several smaller cases that are connected to each other. This way, the similarity computation of the cases will be easier and the reuse of knowledge and experience parts will be possible.

From our perspective many different information can be extracted from one story or interview document:

- Abstract and detailed cases for the case base
- Terms and case structures for the vocabulary
- Similarity measures and relevance information
- Adaptation rules
- Connections between cases

From each interview one or more detailed cases for the case base of a CBR system can be extracted. During an interview the expert speaks about different situations and experiences and for every situation a case can be generated. Based on the generated cases from different interviews, abstract cases can be generated to make several aspects of the interview more comparable and reduce the detailed information to the relevant knowledge. An interview contains the relevant terms of the experience domain and these terms can be used to define and extend the vocabulary and the case structure of a CBR system. Analyzing the occurrences and connections between these terms, similarity measures can be defined and the relevance of terms can be determined. Additionally, based on the comparison of different interviews, adaptation rules and connections between cases could be identified and extracted. The adaptation rules could be used to adapt cases extracted from a specific interview to fit to a new situation that is similar but not the same. The connection between cases will be very helpful, when planning a new Triad talk or moderating one.

Considering all these points, a structural CBR approach would be good for combination with triad talks. The initial knowledge modeling of a structured CBR system is done by an expert of the domain and this can cause a high effort. The methodical listener in a Triad talk writes down the results and stores them in a document. He also could be responsible to transform the document for the use in a structural CBR system. To support the transformation from free text documents into a structural CBR system we developed a framework called FEATURE-TAK, an agent-based Framework for **E**xtraction, **A**nalysis, and **T**ransformation of **U**nstructe**RE**d **T**extual **A**ircraft **K**nowledge. The framework was developed for use in the aircraft domain, but can be used in other domains, too. In the following, we will describe the idea and the individual tasks of the framework.

The framework consists of five components: data layer, agent layer, CBR layer, NLP layer and interface layer. The data layer is responsible for storing the raw data and the processed data for each task. In addition, domain specific information like abbreviations and technical phrases are stored in this layer to be accessible for the other components. The agent layer contains several software agents. For every task an individual agent is responsible. All task agents communicate with a central supervising agent. This supervising agent coordinates the workflow. For visualization and communication purposes for the user, this layer also contains an interface agent. For each task an agent is spawned when starting the framework, but during the workflow additional agents can be spawned to support the initial agents with huge data sets. The NLP layer contains algorithms and methods like part of speech tagging, lemmatization, abbreviation replacement and association rule mining. These algorithms are used by the agents to execute their assigned tasks. The algorithms could either be third party libraries or own implementations. The fourth layer is the CBR layer and is responsible for the communication with a CBR tool like myCBR or jColibri. It contains methods to add keywords to the vocabulary extend similarity measures and generate cases from the input data sets. The last layer contains the graphical user interface of the framework. This user interface can be used to configure the workflow, select input data, and start the workflow. In addition, the user interface presents the results of each task to the user and shows the status of the software agents. In the following the tasks of the framework are described. Figure 3 shows all tasks of the framework FEATURE-TAK.

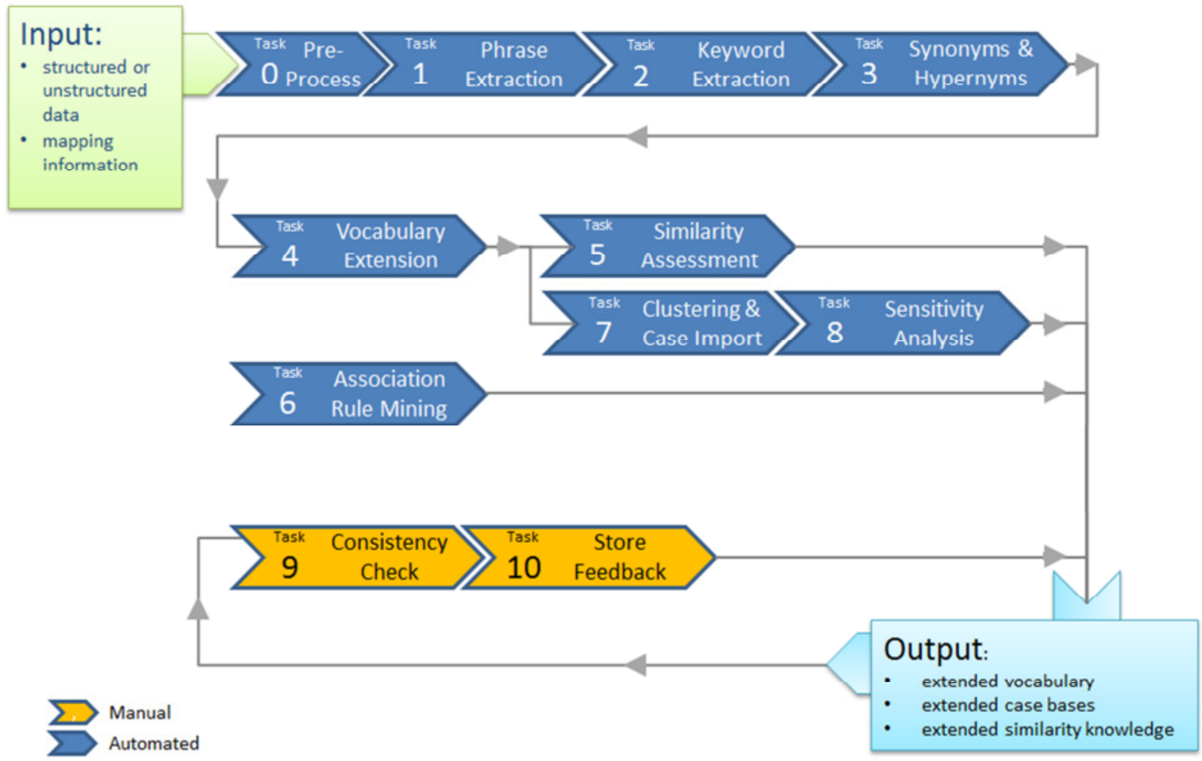

Figure 3: Subtasks of the framework FEATURE-TAK

The first task is the identification and extraction of phrases from the free text components of the input data. The idea is to find recurring combinations of words based on Standard English grammar and domain specific terms. This task has three sub steps: part of speech tagging, multi-word abbreviation identification, and phrase extraction. First, the free text is tagged to identify nouns, verbs, and adjectives. The next step is to identify multi-word abbreviations, because the long form of these abbreviations counts as phrases, too. The last step is to identify phrases based on the tagging information and the word position in a sentence.

The second task is the extraction of keywords from the remaining text and consists also of three sub steps: stop word elimination, lemmatization, and single-word abbreviation replacement. As input for this task, the modified text from task one is used. The stop word elimination is based on Standard English and a white list with words that should not be eliminated. The second sub step identifies abbreviations in the remaining words and replaces them with their long form. For all words the lemmata are determined.

The third task is responsible for identifying synonyms and hypernyms for the extracted keywords and phrases. Therefore, the input for this task is a list of phrases from the first task and list of keywords from the second task. For every keyword the synonyms are identified, based on Standard English and domain specific terms. There are two goals for this task. The first goal is to enrich the vocabulary of our CBR systems and the second goal is to use the synonyms and hypernyms to enhance our similarity measures by extending or generating taxonomies.

This task consists of adding the extracted keywords, phrases, synonyms, and hypernyms to the vocabulary of the CBR systems. The first step is to remove duplicate words and phrases to avoid redundant knowledge. The second step is to check the list of keywords against the list of phrases to identify keywords which occur as phrases. We want to slow down the growth of the vocabulary and therefore we identify keywords that are only occurring as part of a collocation. These keywords are not added to the vocabulary. If a keyword occurs without the context of a collocation, it will be added.

The fifth task is responsible for setting initial similarity values for newly discovered concepts and extends existing similarity measures. The first sub step is to set similarity values between the newly added keywords and phrases and their synonyms. Therefore, the existing similarity matrices are extended and a symmetric similarity is proposed. The value itself could be configured, but we assume an initial similarity for synonyms of 0.8, based on the assumption that the similarity measures can take values from the [0;1] interval. The second step is to use the keywords, phrases, and hypernyms to extend or generate taxonomy similarity measures. The hypernyms serve as inner nodes, while the keywords and the synonyms are the leaf nodes. Keywords and their synonyms are sibling nodes if they have the same hypernym. This second step provides the possibility to model or extend similarity measures based on the layers of a taxonomy and therefore less similarity values have to be set. For values of keywords and phrases that could not be assigned to a taxonomy, no initial similarity value could be set, than 0. To overcome this hurdle, we employ social network analysis (SNA) methods to supplement the similarity between each two values of a given attribute. SNA is based on graph theory and utilizes the structure of the data and the relationships between the different items to reach conclusions about it, and has been used previously to measure the similarity of objects (Ahn, Y., Ahnert, S., et al 2011; Jeh, G., Widom, J 2002).

The sixth task is used to analyze the keywords and phrases and find associations between the occurrence of these words within a data set as well as across data sets. Using association rule mining algorithms like the Apriori (Agrawal Srikant 1994) or the FP-Growth (Borgelt 2005) algorithm, we try to identify reoccurring associations to determine completion rules for our CBR systems to enrich the query.

This task is responsible for generating a case from each input data set and storing it in a case base. To avoid a large case base with hundreds of thousands of cases, we cluster the incoming cases and distribute them to several smaller case bases. Generating an abstract case for each case base, a given query can be compared to the abstract cases and this way a preselection of the required case bases is possible.

In this task the feature weights for the problem description of the given case structure are determined. Not all attributes are equal. In retrieval tasks some attributes are more important to determine which objects are relevant, but how do we identify these attributes, and what is their degree of importance? Can some attributes be detrimental to retrieval? To answer these questions we used sensitivity analysis, and developed a method to calculate a relevance matrix of attributes. A more detailed description of the sensitivity analysis can be found in (Stram Reuss 2016).

Using a framework like FEATURE-TAK can reduce the effort for building and maintaining a CBR system significantly. The framework has been successfully used within the OMAHA project (BMWI 2013) for transforming information from free text documents into a structural CBR system and reduced the effort of modeling the required knowledge.

3 Potential benefits and limits

As described before, the combination of Triad talks and CBR uses the strength of both approaches and tries to minimize the shortcomings. There are several potential benefits for organization using such a combination.

First, the interviews and interview parts become comparable for a computer. It would be no longer necessary to search manually in interview documents, compare documents by hand to find connections or the most relevant information. Second, the reuse of the gained knowledge and experience from Triad talks will be much easier. Using CBR, the effort to find relevant and similar experience to a given situation will be significantly reduced. The retrieval mechanisms of structural CBR systems are very efficient and can handle thousands of cases in several seconds. Third, extracting one or more cases from an interview, distributes the knowledge over several smaller units than a huge text document. Different aspects or stages of an interview could be viewed individually and could be rated with different relevance. This way, the knowledge gained in one interview can be ranked by importance and usability for different situations. Generating abstract cases from the detailed ones, enables a user to compare situations and experiences on higher level and sometimes makes experiences comparable in the first place. Fourth, the combination of Triad talks and CBR enables connections between interviews and interview parts and store these connections. This way, a user would have information about additional cases that may also be useful for a given situation or cases that other users of the CBR system have used in the past to solve a situation. Fifth, the learning capabilities a CBR system can provide. A CBR system can learn new cases, new similarity measures and new adaptation rules. This way, a CBR system can improve over time and become more efficient. Given the feedback of the users of a CBR system to retrieved cases and presented solutions, a CBR system can adapt to new situations.

There are several shortcomings and limits as well, when combining Triad talks and CBR. The first limit is the natural language processing. The interview documents are written by the methodical listener of a Triad talk. Different listeners may use different abbreviations or different terms for the same situation. The syntax of the written free text may not be correct or different languages are used. Most techniques of language processing required correct syntax of sentences and therefore the results from analyzing free text may vary. A second limit is initial knowledge modeling that is required for a CBR system. This effort can be very high, depending on the domain. Especially, modeling the similarity measures requires expert knowledge of the used domain. With frameworks like FEATURE-TAK, the effort can be reduced, but a still an expert is required to validate the results of the Framework. Third, the integration of the CBR system into the interview planning and the interview itself to support the methodical listener requires the adaptation of the knowledge management process in an organization. The CBR system should be used before and during the interview and a process for to manage the quality of the knowledge in the CBR system may also be necessary.

4 Summary and outlook

In our article we described a combination approach between the knowledge management methods Triad talks and Case-Based Reasoning. We gave on overview of our combination idea and how the knowledge of both methods can be combined to create synergies. In addition, we described our framework FEATURE-TAK, which supports the transformation of knowledge between Triad talks and CBR systems and described the potential benefits and limitations of our combination approach.

We have analyzed exemplary Triad talks and developed an initial methodology to transform the knowledge from an interview into vocabulary, case structure, similarity measures, and cases. Our next steps will be to extend FEATURE-TAK to a more generic framework and realize interface for the domain specific knowledge. After that we will be able to test our combination approach and knowledge transformation from Triad talks into CBR systems in different domains. In addition, we will extend our framework with learning mechanisms to learn from user feedback.

Literature

Aamodt A, Plaza E (1994) Case-Based Reasoning: Foundational Issues, Methodological Variations and System Approaches, IN: AI Communications 7, Nr.1, pp 39-59, 1994

Agrawal, R, Srikant, R (1994) Fast algorithms for mining association rules in large databases. In: Proceedings of the 20th International Conference on Very Large Data Bases. pp. 487-499. VLDB '94, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA 1994

Aha D, Breslow, L et al (2001) Conversational Case-Based Reasoning IN: Applied Intelligence 14, pp 9-32, 2001

Ahn, Y, Ahnert, S, et al (2011) Flavor network and the principles of food pairing, Scientific reports 1 2011

BMWI (2013) Luftfahrtforschungsprogramms V (2013), http://www.bmwi.de/BMWi/Redaktion/PDF/B/bekanntmachung-luftfahrtforschungsprogramm-5,property=pdf,bereich=bmwi2012,sprache=de,rwb=true.pdf

Borgelt, C (2005) An implementation of the FP-growth algorithm. In: Proceedings of the 1st International Workshop on Open Source Data Mining: Frequent Pattern Mining Implementations. pp. 1-5. OSDM '05, ACM, New York, NY, USA 2005

Dick, Michael (2006) Triadengespräche als Methode der Wissenstransformation in Organisationen. IN: V. Luif, G. Thoma & B. Boothe (Hrsg.). Beschreiben – Erschliessen – Erläutern. Psychotherapieforschung als qualitative Wissenschaft, Pabst Verlag 2006

Jeh, G, Widom, J (2002) Simrank: a measure of structural-context similarity. IN: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 538-543 2002

Kolodner, Janet (1993) Case-Based Reasoning, Morgan Kaufmann Publishers Inc, San Mateo 1993

Lenz, Mario (1999) Case Retrieval Nets as a Model for Building Flexible Information Systems, Humboldt-Universität Berlin, Mathematisch-Naturwissenschaftliche Fakultät II, Dissertation, 1999

Pfuhl, Markus (2003) Case-Based Reasoning auf der Grundlage Relationaler Datenbanken, 1. Auflage, Deutscher Universitätsverlag/GWV Fachverlage GmbH Wiesbaden, 2003

Reinmann-Rothmeier G, Erlach C. et al (2000) Erfahrungsgeschichten durch Story-Telling: Eine multifunktionale Wissensmanagement-Methode, Forschungsbericht Nr.127, Ludwig-Maximilians-Universität, München 2000

Richter, Michael (2003) Handbuch der künstlichen Intelligenz, Kapitel Fallbasiertes Schließen, pp 407 – 430, Oldenbourg Wissenschaftsverlag 2003

Stary C, Maroscher M et al (2013) Wissensmanagement in der Praxis: Methoden, Werkzeuge, Beispiele, Carl Hanser Verlag GmbH, München 2013

Stram, R, Reuss, P, et al (2016) Relevance matrix generation using sensitivity analysis in a case-based reasoning environment. In: Proceedings of the 25th International Conference on Case-based Reasoning, ICCBR 2016. Springer Verlag 2016

Address of authors

M.Sc. Pascal Reuss
Universität Hildesheim
Intelligente Informationssysteme
Universitätsplatz 1
D-31141 Hildesheim
Email: reusspa@uni-hildesheim.de

Prof. Dr. Michael Dick
Dipl.-Päd. Wilhelm Termath
Otto-von-Guericke Universität Magdeburg
Fakultät für Humanwissenschaften
Institut I: Bildung, Beruf und Medien
Zschokkestr. 32
D-39104 Magdeburg

Prof. Dr. Klaus-Dieter Althoff
Universität Hildesheim – Intelligente Informationssysteme
Deutsches Forschungszentrum für Künstliche Intelligenz
Kompetenzzentrum Case-Based Reasoning
Trippstadter Straße 122
D-67663 Kaiserslautern