Mapping and Usage of Know-How Contributions

Arnon Sturm[1,2(✉)], Daniel Gross[1], Jian Wang[1,3],
Soroosh Nalchigar[1], and Eric Yu[1]

[1] University of Toronto, Toronto, Canada
{daniel.gross,eric.yu}@utoronto.ca,
soroosh@cs.toronto.edu
[2] Ben-Gurion University of the Negev, Beer–Sheva, Israel
sturm@bgu.ac.il
[3] Wuhan University, Wuhan, China
jianwang@whu.edu.cn

Abstract. Mapping know-how, which is knowledge of how to achieve specific goals, is important as the creation pace and amount of knowledge is tremendously increasing. Thus, such knowledge needs to be managed to better understand tradeoffs among solutions and identify knowledge gaps. Drawing from goal-oriented requirements engineering, in this paper we propose a specialized (and light weight) use of concept maps to map out contributions to problem-solving knowledge in specific domains. In particular, we leverage on the means-end relationship which plays a major role in such domains and further extend it to be able to depict alternatives and tradeoffs among possible solutions. We illustrate the approach using problems and solutions drawn from two domains and discuss the usefulness and usability of the know-how maps. The proposed mapping approach allows for a condensed representation of the knowledge within a domain including the contributions made and the open challenges.

Keywords: Knowledge mapping · Requirement engineering · Concept map

1 Introduction

The state of the art in fields is a fast moving target. With innovation occurring globally at a fast pace, researchers and practitioners who are pushing the boundaries to better deal with new problems and needs, expend significant efforts to keep up with the current state of the art. To stay up-to-date and to make new contributions to bodies of knowledge, researchers and practitioners must continuously maintain an overview of a field, understand the problems addressed and solutions proposed, as well as identify the outstanding issues that should receive further attention. Given the fast pace of new developments, keeping such an overview up-to-date is challenging. Researchers and practitioners typically make use of, or produce, literature reviews to better understand and map out specific domains. Some use informal literature maps to map out fields of research, indicating with overlapping circles, or boxes with lines in between them,

sub-topics or themes and relationships between them. Some also use their own accumulated list of scholarly references typically included in bibliographic software, as their main personal "knowledge" database which is then tagged and labeled to cue of the existing and expanding works in their fields of interest.

While such approaches are commonly used, they much depend on the researchers and practitioners capability to identify means-end relationships details, which are not guided by the following questions: how problems are characterized, what problems are already addressed, how proposed solutions are making significant contributions to problems, what proposed solutions fail to address, and where further contributions and improvements can be sought.

We observe that problems and solutions in fields of study, and the contribution relationships between problems and solutions can be characterized as *means* that come to address specific *ends* in some better way. We observed a lack of exploiting such a conceptual relationship in current approaches to mapping out engineering domains. Means-ends relationships that are made use of relationships between problems and solutions are such conceptual relationships that are used only to a limited extent in the context of knowledge mapping. Nevertheless, these are widely used within goal-oriented requirements engineering (GORE) approaches [27], in general, and the i* (pronounced i-star) goal-oriented modeling framework approach, in particular [27]. The i* approach, for example, has at its center the means-ends relationship, and the capability to differentiate alternate means towards some end by indicating their differing contributions towards desired quality objectives (by use of additional contribution links). Following the i* notions and based on [9, 23], we propose a knowledge mapping approach to represent and map out problems and solutions in domains which relies on the means-end relationships. We envision that using such an approach would better support researchers and practitioners in representing, capturing and reasoning about research advancements in such domains.

In this paper, we describe a means-end oriented approach for knowledge mapping, and illustrate its use for various domains of technology-oriented (such as the data mining domain) and business-oriented (such as the customer relationship management domain). In addition, as we are aiming at mapping out relationships among domains we further define the cross relationships among domain maps. In particular, we demonstrate how the data mining domain is used to solve problems in the domain of Customer Relationship Management (CRM), which can indicate ways communities working in those different domains could collaborate to better solve existing problems. Furthermore, in this paper we stress the notion of context and quantitative contribution. In addition, we present an initial evaluation of the approach and further discuss our vision of future developments along the lines of the proposed approach.

The paper is organized as follows. Section 2 discusses related work. Section 3 introduces the mapping approach and demonstrates it using examples from the two domains. Section 4 introduces the evaluations we performed to assess the usefulness of the proposed approach. Section 5 further discusses the proposed approach. Finally, Sect. 6 concludes and further elaborates on our vision regarding the usage of such knowledge maps in the future.

2   Related Work

To date, little work has been done to offer a conceptual approach to mapping out domain knowledge within one or more fields. Researchers mainly use literature reviews, including systematic reviews [14], tagging and classifications approaches[1] to accumulate and organize literature in one or more domains [16]. Such approaches help cluster literature along themes or viewpoints of interest, however, offer little insights into the structure of domain knowledge itself, such as to characterize problems, solutions and innovative contributions. Some works aim at mapping out inter-disciplinary research using approaches to visualizing domains knowledge and kinds of relationships between domains knowledge. This includes, for example, citation graphs, subject heading and terminology clustering [4, 5, 17], as well as work toward output indicators for interdisciplinary science and engineering research, which give insight into social dynamics that lead to knowledge integration [24].

While these approaches offer important sign-posts for researchers to orient themselves in disciplinary and cross-disciplinary fields of research, such approaches typically use quantitative measures and descriptive statistic approaches, including novel multi-dimensional networking graphics, to indicate citation structures and interconnections between topics across disciplines, without, however relating the conceptual structuring of the knowledge and evolution of such knowledge in domains.

Some research has been done to offer a conceptual view of literature in a domain and to support a conceptual consolidating of scholarly works. This includes concept maps [19], cause maps [7], and claim-oriented argumentation [21, 22]. One particularly noteworthy line of research is VIVO, a semantic approach to scholarly networking and discovery. VIVO provides a semantic-web based infrastructure and ontologies to represent, capture and make discoverable conceptual linkages that define scholars, and their scholarly work [2].

While these offer useful structuring of domain knowledge, they do not specifically focus on linking problem and solution knowledge within and across domains, in general, and during knowledge innovation and evolution, in particular.

As researchers and practitioners are looking for innovation, they would benefit from supporting tools that would help them in clustering related topics (clustering), as well as explicate problems and solutions (expressiveness), to represent and reason about differences between existing solutions (reasoning), and to be able to adjust the mapping as the domain evolves over time (dynamic evolution). Indeed, the aforementioned approaches offer different kinds of textual, conceptual and visual mapping over domains. However, as mentioned, they lack essential capabilities to evaluate or compare state-of-the-art studies at an engineering knowledge level. Table 1 presents a comparison of the approaches with respect to the needed capabilities.

In addition to mapping specific domains, both researchers and practitioners have a need to explore the possibilities of adopting or adjusting existing solution in one

---

[1] There are also reference management systems like EndNote and Mendeley that support classification using folders and tagging.

Table 1. Comparing mapping techniques

Approach	Advantages	Drawbacks
Literature survey	Literature review enables clustering and since it uses natural language its expressiveness is extensive	As a literature review provides mapping by textual means, a direct reasoning on it is challenging. Furthermore, as such review is done once in a while its relevance is limited. In addition, usually such reviews are taking an a priory point of view and thus might be leaning towards specific directions
Classification	Classification aims at clustering existing work and also facilitates the addition of related knowledge into existing clusters	Classification has limited expressiveness as it indicates only the association to specific categories and not the reason for that. Also, classification does not allow for clear differentiation among works within each category. Usually, classification is done at high level of abstraction and thus provides limited reasoning capabilities. Also, changing clusters requires re-assignment of the already categories work, so the support for dynamic evaluation is limited
Cause map	Cause maps do provide some kind of clustering, though it is not explicated. It allows for some reasoning on various alternatives to achieve an end and also support dynamic evolution of domains as alternatives can be easily be incorporated in such map	Cause maps focuses on means-ends relationships in a conservative manner, usually, without explicit differentiation among alternatives to achieve an end
Concept maps	Concept map has an extensive expressiveness as they can be used to specify any kind of claims	Concept maps are similar to literature review, but add the identification of key concepts. Thus, reasoning remains a challenge. Clustering also should be explicated as it is not clear in such maps
Argumentation	Argumentation approaches can be easily be equipped with extended expressiveness. Being formal in nature reasoning with such approaches is easy. Also, by various arguments differentiations among various works can be explicated	By simply looking at the set of arguments it would be difficult to identify clustering. Yet, this can be achieved through a reasoning procedure

domain to problems in other domains. This exploration would increase collaboration among different research communities.

3 Specializing Concept Maps for Specific Domains

In this section we introduce an approach to map out domains that are characterized by the means-end relationship. To address this need we extend concept map by the means-end relationships to enable reasoning and evolution of existing maps. We begin with briefly presenting the domains that are used in this paper for demonstrating the approach. Then, we present the concepts used within the approach and demonstrate these on the introduced domains.

3.1 Two Examples of Know-How Domains

In this paper, we refer to two domains, namely, the business domain of customer relationship management and the technical domain of data mining. For each domain, some of the content and problems they address are briefly discussed.

3.1.1 The Customer Relationship Management Domain

Customer Relationship Management (CRM) is a comprehensive process of building long term and profitable relationships with customers. It includes four dimensions: customer identification, customer attraction, customer retention, and customer development [16]. Customer retention aims to keep loyal customers by reducing their defections. To this end, approaches to predicting customer churn were developed [11, 23]. Churn prediction is a well-studied domain with various proposed techniques. Moreover, various experiments have been undertaken to compare those techniques (e.g., [26]). Those experiments further asked the following questions: How to examine the existing techniques and to identify opportunities for new contributions in that specific domain for improving customer relationship management practices? How to find techniques from other disciplines and use them for churn prediction? We show how the knowledge mapping approach helps in answering such questions.

3.1.2 The Data Mining Domain

Data mining, the process of extracting interesting patterns and trends from large amounts of data, has gained much attention and success in many scientific and business areas [9]. The main goal is to build a model from data [7]. Major classes of data mining tasks are: predictive modeling, clustering, pattern discovery, and probability distribution estimation [18]. Predictive modeling deals with the problem of performing induction over the dataset in order to make predictions. One of the ways of predictive modeling is classification which builds a concise model or classifier that represents distribution of class labels in terms of predictor features.

Data mining is a domain of rapidly growing interest in which various algorithms and techniques are being proposed. For example, to support data classification different algorithms were proposed, including Support Vector Machines (SVM), Naïve Bayes,

and Decision Trees. Also for the other data mining tasks, several algorithms exist and new ones appear continually, making it hard to keep track over time, as well as choose one over another. This problem is further complicated by the fact that various algorithms have different performance rates over different datasets and there is no single algorithms that outperforms all the others for all datasets. Furthermore, even specialists lack comprehensive knowledge about the full range of algorithms and techniques along with their performances. As a result, understanding the strengths and weaknesses of various algorithms and choosing the right one is a challenging task and a critical step both during the design and implementation of data mining applications, as well as when researching novel solutions. In the next sections we illustrate how a knowledge mapping approach can partially alleviate these challenges.

3.2   The Knowledge Mapping Approach

This section presents an approach to conceptually map domains that has at its center means-end relationships. We applied the approach to map out portions of engineering domains including agent-oriented software engineering, geo-engineering, web mining, and documentation of software architecture. We adopted a minimal set of modeling constructs to two types of nodes and a number of types of links. By convention, the map is laid out with problems or objectives at the top and solutions at the bottom.

Figures 1 and 2, respectively illustrate parts of the knowledge maps of the data mining and the customer relationship management domain. Note the maps purpose is to illustrate the concepts used in the proposed mapping approach rather to demonstrate comprehensive maps. For our modeling needs we used the concept mapping toolset
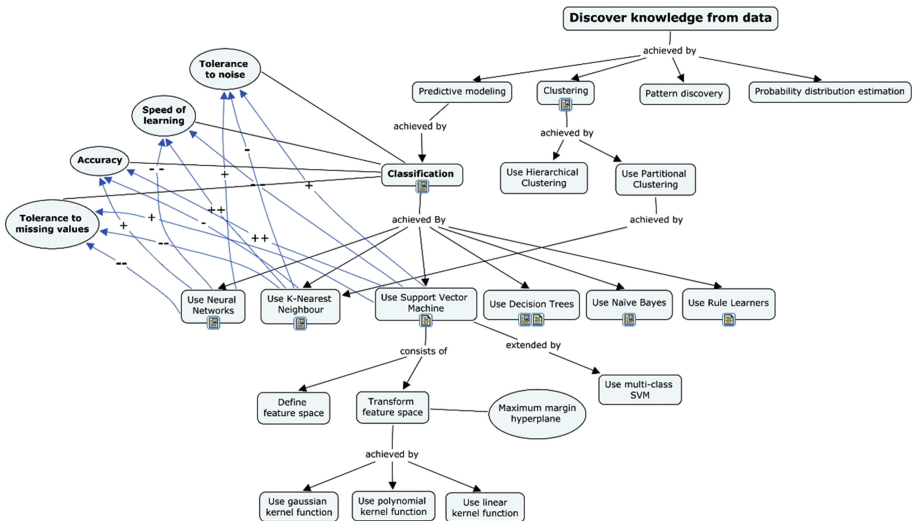

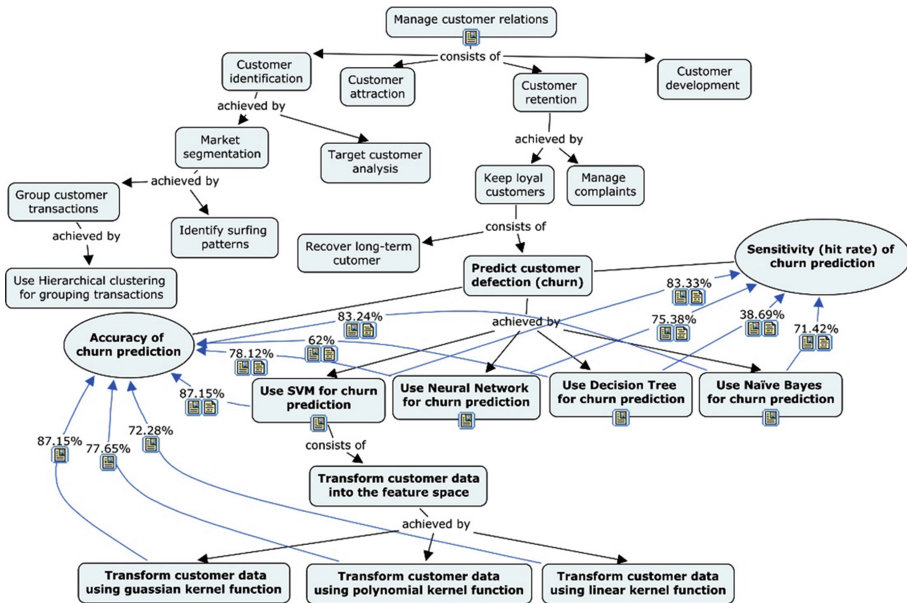**Fig. 1.** A partial knowledge map of the data mining domain

Fig. 2. A partial knowledge map of the customer relation management domain

cmap tools[2] to draw the knowledge maps. Using the cmap tool allows us to benefit from all implemented features of the platform, including collaborative modeling and sharing of concept maps. It also provides a modeling mechanism to support the connections among several domains.

In the following we explain and demonstrate the constructs used in the proposed mapping approach.

- The **task** is the main element used to define means-ends relationship, and when chained together, means-ends hierarchies. A task can be interpreted either as a problem or a solution. Typically, it is named with a verb phrase, and is graphically depicted as a rectangular shape with rounded corners. For example, in Fig. 1, the task "*Discover knowledge from data*" is a typical problem in the data mining domain that needs to be addressed. It can be addressed by tasks such as "*Predictive modeling*", "*Clustering*", "*Pattern discovery*", and "*Probability distribution estimation*". Each of these solutions can in turn be viewed as sub-problems that need further addressing. For example, the "*Predictive modeling*" problem can be addressed by "*Classification*".
- A **quality** element is used to express quality attributes that are desired to sufficiently hold when addressing tasks. A quality is depicted as an ellipse, and is typically named with adverbial or adjectival phrases or quality nouns (e.g., "-ilities").

Example qualities in Fig. 1 are "*Accuracy*", "*Tolerance to noise*", and "*Speed of learning*".

- Links connect tasks and qualities. We propose the use of following link types:
  - The **achieved by** link represents a means-end relationship. The arrow points from the "end" to the "means". Figure 1 indicates that "*Classification*" is one way to achieve "*Predictive modeling*". "*Use Neural Networks*", "*Use K-Nearest Neighbour*", "*Use Support Vector Machine*", "*Use Decision Trees*", "*Use Naive Bayes*" and "*Use Rule Learners*" are alternative ways of achieving "*Classification*". While in Fig. 2, "*Keep loyal customers*" and "*Manage complaints*" are solutions to achieve the task (in this case the problem of) "*Customer retention*".
  - The **consists of** link indicates that a task has several sub-parts, all of which should be addressed for the parent task to be addressed. In Fig. 1, "*Use Support Vector Machine*" consists of "*Define feature space*", and "*Transform feature space*", among other problems that need to be addressed. While in Fig. 2, "*Manage customer relations*" consists of "*Customer identification*", "*Customer attraction*", "*Customer retention*", and "*Customer development*".
  - The **association link** (an unlabeled and non-directional link) indicates the desirable qualities that should sufficiently hold for a given task, once addressed. These qualities are later also to be taken into account when evaluating alternative ways for addressing the task. For example, in Fig. 1 "*Accuracy*" and "*Speed of learning*" are qualities that could serve as criteria when evaluating different ways to address "*Classification*". In Fig. 2, "*Sensitivity (hit rate) of churn prediction*" and "*Accuracy of churn prediction*" are two qualities to evaluate different solutions for task "*Predict customer defection (churn)*".
  - The **extended by** link indicates that the target task is an extension of the source task. For example, in Fig. 1, "*Use multi-class SVM*" is an extension of "*Use Support Vector Machine*". All qualities that hold for the parent task also hold for its extensions.
  - The **contribution link** (a curved arrow) indicates a contribution towards a quality, which can be directed either from a task or another quality. Following the i* guidelines, the contribution is subjective and can range from positive to negative contribution. For example, in Fig. 1, "*Use Neural Networks*" contributes positively ("+") to "*Accuracy*" and "*Tolerance to noise*", but negatively ("−") to "*Tolerance to missing values*" and "*Speed of learning*". However, in the case of the knowledge map, where an objective scale measure can be associated with a solution that addresses a particular quality, these measures can also be associated with the link. For example, as shown in Fig. 2, the "*accuracy of churn prediction*", according to [28] if it is addressed by "*Use SVM for churn prediction*" is 87.15 %, by "*Use Neural Network for churn prediction*" is 78.12 %, by "*Use Decision Tree for churn prediction*" is 62 %, and, finally, by "*Use Naive Bayes for churn prediction*" is 83.24 %.

Each element in the knowledge map can have a context associated, such as a conditions, datasets, experimental settings, and so on, and must have a set of references,

which are the knowledge sources. These help justifying the existence of the element within the map. To avoid cluttering we have omitted such contexts or references elements in Figs. 1 and 2. Instead we included them in the attached note icons and reference icons of the tasks in the figures, which are displayed when clicking the icons in the tool. For example, the attached reference icon of "*Classification*" in Fig. 1 shows that the knowledge source of the task is from [15]; the attached note icon of the number "78.12 %" in Fig. 2 shows that the experiment is conducted on a subscriber dataset consisting of 100,000 customers with 171 potential predictor variable, and the attached reference icon shows that the experiment is reported in [28].

It is important to note that a map is essentially an index to the actual knowledge. The purpose of a map is not to represent the entire knowledge but rather organize the knowledge to increase its accessibility.

To construct the knowledge map in Fig. 1, we referred to the definitions of data mining, classification and support vector machine in Wikipedia and analyzed a survey paper on classification techniques [15], as well as a tutorial paper on support vector machines [3]. Similarly, to construct the knowledge map in Fig. 2, we analyzed a survey paper on the application of data mining to customer relationship management [18], as well as a paper on customer churn prediction using SVM [28]. Following these resources, we were able to construct the map while having supporting evidences for the claims implied by nodes and links included in the map.

As we further identify innovation in a domain (by identifying papers reporting the innovation), we make additions to the maps. For example, Fig. 2 shows results of our analysis of another paper on customer churn prediction based on SVM [25], which compares several classification techniques by considering not only the accuracy which is compared in [28], but also the sensitivity (hit rate) of churn prediction. Note that the work of Xia and Jin [25] adopted a different dataset (the machine learning UCI database of University of California). We then further analyzed a survey paper on clustering [12] and a paper on grouping customer transactions based on hierarchical pattern-based clustering [26]. All the model elements and the links in the knowledge map are derived from contents reported in the above mentioned knowledge sources. However, due to space limitations, we only show a small part of the knowledge on clustering in Figs. 1 and 2.

To demonstrate the mapping work procedure, in the following we use Fig. 2 to illustrate the construction process of a knowledge map. The first step is to identify tasks. For example, in Fig. 2, the problems of the CRM dimensions, the CRM elements, and the solutions such as data mining functions and specific data mining techniques, are identified and recorded as tasks. Afterwards, we link the identified tasks by *consists-of* or *achieved-by* links according to their relations. For example, the CRM dimensions such as "*Customer identification*" and "*Customer attraction*" are linked with "*Manage customer relations*" by *consists-of* links.

The next step is to identify qualities related to tasks and use association links to associate them with their respective tasks. For example, as shown in Fig. 2, the "*Accuracy of churn prediction*", a criteria used in the comparison among different solutions, is identified as a quality of the task "*Predict customer defection (churn)*". Finally, we need to link the alternative solutions with qualities by contribution links. In Fig. 2, the numbers on the contribution links originated from the comparison tables in [28].
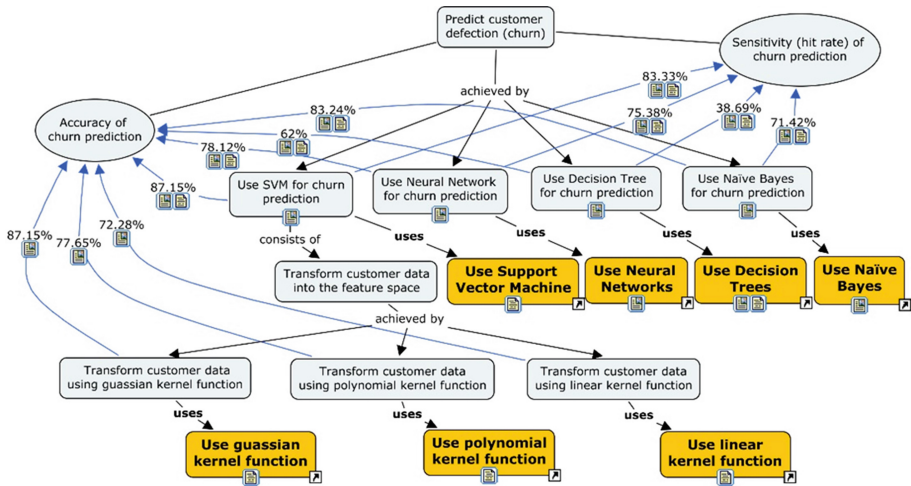
Fig. 3. Connecting the knowledge of the two domains

As noted, one of the approach objectives is to reveal and map out relationships across different but related domains. Some technologies in one domain may be applied in other domains. For example, techniques in the data mining domain have been applied in the CRM domain. Since a knowledge map mainly embodies the knowledge within a specific domain, we introduce a new link type named "uses" to connect tasks in different domains. For example Fig. 3 illustrates a connection between the knowledge in data mining domain and the CRM domain, by linking the task "*Use SVM for churn prediction*" via a *uses* link to the technology "*Use Support Vector Machine*" of the data mining domain included in Fig. 1. In this case, the CRM domain is named as the problem domain while the data mining domain is the solution domain. Including the knowledge of one domainin in another domain can contribute to discover heuristic solutions for the problems in other domains.

Referring to the questions raised in Sect. 3.1, the knowledge map in Fig. 2 can be used to map out the existing techniques in the CRM domain. Furthermore, it is easy to the add new contributions to the existing knowledge map based on the problems they address and compare them with existing contributions using the contribution links. New techniques from other disciplines can be identified from the "uses" links. As illustrated in Fig. 3, considering such links can assist in finding appropriate solutions by leveraging the knowledge in other domains. These can help answer questions such as what techniques from other disciplines have been used for churn prediction. Also, our approach can help users to examine and compare existing techniques and find opportunities for new contributions. For example, if the "speed of learning" is a major concern for the churn prediction in a certain context, Fig. 1 can facilitate identifying that "Use K-Nearest Neighbour" will be the best choice for the problem at hand. Finally, regarding the question about selecting the right data mining algorithm raised in

Sect. 3.1.2, we believe that adding contexts and references for contribution links can help gaining a better understanding about the comparisons of various algorithms.

4   Evaluation of the Know-How Mapping Approach

In order to evaluate the approach, we have performed several evaluation steps during its development. The evaluation consists of two main themes: comprehension of maps and construction of maps.

For the comprehension theme we compared the understanding of a knowhow map versus a literature review[1]. In that evaluation, we used a preliminary notation which was a sub-set of the i* framework (that represents tasks and qualities along with the related relationships). We had twelve subjects of which four were familiar with i* and eight which were not familiar with that framework. The four subjects who were familiar with i* got the knowledge map of a web mining domain along with other four subjects. The other four subjects got the literature review of the same domain. We further made sure that information in both the map and the literature is equivalent. Upon getting the domain knowledge (either as a map or as a text) we ask the subjects to answer a questions related to problems, solutions, properties, and tradeoffs in the domain at hand. The results indicated that having the knowledge map better allowed the subjects to understand the domain and in less time than was required by the literature review, even by those subjects who were unfamiliar with i*.

For the construction theme, we recruited four graduate students and after training them with the proposed approach, we ask them to map out their own research domain. The mapping was performed in a few stages, so we were able to control and give feedback on their mapping. We then the reviewed the resulted maps and ask them to fill out a questionnaire indicating the usability of the proposed approach. Analyzing their responses we concluded the following:

- The approach is easy to use for the purpose of mapping a literature review.
- The approach helps in organizing the knowledge in a way that facilitates grouping of similar studies, as well as, differentiating among related studies.
- The approach facilitates identification of research gaps and possible contributions from other domains.
- The approach helps in positioning own research.
- The approach encourages critical thinking with respect to literature reviews.

Although further validation is required, the results obtained so far indicated that the approach does provide meaningful benefits.

5   Discussion

Using the specialized concept maps to connect between a conceptual representation of problems and solutions in domains of interest supports researchers and practitioners in quickly gaining insights into the problems they deal with and solution practices available to them, within and across domains. A key advantage that such an overview

offers is the systematic overview of solution approaches that could fit problems thereby reducing the risk of missing relevant techniques to address specific challenges. However, while the proposed approach facilitates representing problems and solutions in existing state-of-the-art, we encountered a number of challenges:

Conceptual Mismatch: Identifying problems, solutions, qualities, and the relationships among them is often non-trivial. Researchers and stakeholders often present needs and benefits in solution-oriented terminology and languages and neglect the connection with the problem-oriented aspects.

Naming Decompositions: During the construction of a knowledge map elements are decomposed into lower level elements. Decomposition is the main mechanism to unearth variation and differences in approach details (solution features) that matter with respect to qualities. However, in some domains it appears difficult to identify and name those solution feature "components" that differentiate among alternative approaches. This suggests that more holistic representations of solution approaches, or, finer-grained concept map based analysis guidelines are needed to help make explicit in what way proposed solutions differ in their details.

Multiple Vantage Points and Terminology Use: Because of different viewpoints map creators might take, they may develop maps differently, both in terminology and in the abstraction level. Furthermore, it is in the purview of the map creator to decide which level of abstraction is the most fitting to express problems and solution approaches. When constructing larger maps out of contributions from different map authors, aligning the levels of abstraction is non-trivial.

Scalable Tool Support: Better tool support is needed. Using concept maps we took advantage of existing tools, and their "scalability" features such as: element expanding/ collapsing and map referencing.

Domain Knowledge Extraction: Currently, knowledge extraction and its mapping are done manually. This introduces a burden on adopting the approach. Nevertheless, we envision crowd-mapping as an approach that distributes the burden across interested participants, who benefit from mutual contributions, and approaches to automated concept extraction from bodies of text guided by the proposed concepts that link needs with solutions.

6 Conclusion and Future Work

We propose an approach to map out problem-solution oriented fields using a light-weight modeling technique, based on concepts borrowed from the area of goal-oriented requirements engineering. We argue for benefits that such an approach would offer, such as the ability to represent and facilitate the analysis for novel solution approaches in light of their quality properties and to identify gaps of un-addressed problems. We also illustrated the ability to represent the use of solutions drawn from more than one domain, and how these contribute to improve the ability to address problems at hand, whilst also having relevant problem qualities in mind. We believe that the approach is applicable to any domains which aim to identify better solutions to well defined

problems, and hence its characterizations fits with the problem-solution means-end chains the proposed approach represents. We also note that benefits for such domains vary and depend on the domain maturity, such as, whether problems are already well understood and solutions already worked out. In particular, the approach works best in cases where domains are mature enough and a large body of knowledge and terminology has been established. In such cases, the mapping would be helped by existing domain resources (i.e., the research literature, such as papers and textbooks), which would likely already have established a common and unified terminology. On the other hand, domains which are evolving would probably use various sets of terminology which would make the mapping difficult. To further explore and facilitate the use of knowledge mapping we plan to expand knowledge map capabilities in a number of directions. We aim to further develop guidelines for map creators to support extracting knowledge from research domains and including them in knowledge maps; to support scalability by developing a framework for mapping and searching knowledge maps; to support a crowd-mapping approach where different stakeholders contribute to creating, arguing about and improving a collaborative created knowledge map; to support for trust mechanisms, as well as, evidence based augmentations of knowledge maps that offer further validity insights; to develop semi-automated reasoning support to identify gaps or even possible solution approaches to already identified gaps, with searches across different knowledge maps; and develop automated extraction of knowledge mappings from bodies of engineering texts, guided by core concepts proposed in this paper. We are also planning further evaluations for testing the benefits of the proposed approach.

References

1. Abrishamkar, S.: Goal-Oriented Know-How Mapping- Modelling Process Documentation, a Prototype, and Empirical Studies, M.I. thesis, University of Toronto (2013)
2. Börner, K., Conlon, M., Corson-Rikert, J., Ding, Y. (eds.): VIVO: A Semantic Approach to Scholarly Networking and Discovery. Morgan & Claypool Publishers LLC, San Rafael (2012)
3. Burges, C.J.C.: A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Disc. **2**(2), 121–167 (1998)
4. Chen, C.M., Paul, R.J.: Visualizing a knowledge domain's intellectual structure. Comput. **34**(3), 65–71 (2001)
5. Chen, C.M., Paul, R.J., O'Keefe, B.: Fitting the jigsaw of citation: information visualization in domain analysis. J. Am. Soc. Inform. Sci. Technol. **52**(4), 315–330 (2001)
6. Eden, C., Ackermann, F., Cropper, S.: The analysis of cause maps. J. Manage. Stud. **29**(3), 309–324 (1992)
7. Giraud-Carrier, C., Povel, O.: Characterising data mining software. Intell. Data Anal. **7**(3), 181–192 (2003)
8. Gross, D., Sturm, A., Yu, E.: Towards know-how mapping using goal modeling. In: International iStar Workshop, pp. 115–120 (2013)

9. Han, J., Kamber, M., Pei, J.: Data Mining: Concepts and Techniques. Data Management Systems, 3rd edn. Morgan Kaufmann, Waltham (2012)

10. Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: a review. ACM Comput. Surv. (CSUR) **31**(3), 264–323 (1999)

11. Kawale, J., Pal, A., Srivastava, J.: Churn prediction in MMORPGs: a social influence based approach. In: International Conference on Computational Science and Engineering, vol. 4, pp. 423–428 (2009)

12. Kitchenham, B., Brereton, P., Budgen, D., Turner, M., Bailey, J., Linkman, S.: Systematic literature reviews in software engineering - a systematic literature review. Inf. Softw. Technol. **51**(1), 7–15 (2009)

13. Kotsiantis, S.B.: Supervised machine learning: a review of classification techniques. Informatica **31**, 249–268 (2007)

14. Kwasnik, B.: The role of classification in knowledge representation and discovery. Libr. Trends **48**, 22–47 (1999)

15. Leydesdorff, L., Rafols, I.: A global map of science based on the ISI subject categories. J. Am. Soc. Inf. Sci. Technol. **60**(2), 348–362 (2009)

16. Ngai, E.W.T., Xiu, L., Chau, D.C.K.: Application of data mining techniques in customer relationship management: a literature review and classification. Expert Syst. Appl. **36**, 2592–2602 (2009)

17. Novak, J. D., Cañas, Al. J.: The Theory Underlying Concept Maps and How To Construct and Use Them, Institute for Human and Machine Cognition (2006)

18. Panov, P., Soldatova, L.N., Džeroski, S.: Towards an ontology of data mining investigations. In: Gama, J., Costa, V.S., Jorge, A.M., Brazdil, P.B. (eds.) DS 2009. LNCS (LNAI), vol. 5808, pp. 257–271. Springer, Heidelberg (2009)

19. Shum, S.B., Motta, E., Domingue, J.: ScholOnto: An ontology-based digital library server for research documents and discourse. Int. J. Digit. Libr. **3**(3), 237–248 (2000)

20. Simsek, D., Buckingham Shum, S., Sandor, A., De Liddo, A., Ferguson, R.: XIP Dashboard: visual analytics from automated rhetorical parsing of scientific metadiscourse. In: Proceedings of 1st International Workshop on Discourse-Centric Learning Analytics (2013)

21. Sturm, A., Gross, D., Wang, J., Yu, E.: Analyzing engineering contributions using a specialized concept map. In: CAiSE (Forum/Doctoral Consortium), pp. 89–96 (2014)

22. Wagner, C.S., Roessner, J.D., Bobb, K., Klein, J.T., Boyack, K.W., Keyton, J., Rafols, I., Börner, K.: Approaches to understanding and measuring interdisciplinary scientific research (IDR): a review of the literature. J. Informetr. **5**(1), 14–26 (2001)

23. Xia, G., Jin, W.: Model of customer churn prediction on support vector machine. Syst. Eng. Theory Pract. **28**(1), 71–77 (2008)

24. Yang, Y., Padmanabhan, B.: GHIC: A hierarchical pattern-based clustering algorithm for grouping Web transactions. IEEE Trans. Knowl. Data Eng. **17**(9), 1300–1304 (2005)

25. Yu, E., Giorgini, P., Maiden, N., Mylopoulos, J.: Social Modeling for Requirements Engineering. MIT Press, Cambridge (2011)

26. Zhao, Y., Li, B., Li, X., Liu, W., Ren, S.: Customer churn prediction using improved one-class support vector machine. In: Li, X., Wang, S., Dong, Z.Y. (eds.) ADMA 2005. LNCS (LNAI), vol. 3584, pp. 300–306. Springer, Heidelberg (2005)