Fostering Learning Communities based on Task Context

Niels PINKWART
University of Duisburg-Essen, Germany

Abstract. This paper presents an approach to establish and support learning communities. Based on task context information (which is extracted from multiple sources) and relying on the documents users created as primary source of information, the concept of a "peer recommender system" is presented which internally makes use of a mixture of different similarity measures – including, e.g., archive distance measurements and ontology based techniques.

1. Introduction

In recent years, the use of advanced computing and information processing techniques in education has significantly increased. Immediate benefits of digitally available resources come with the options of re-use and sharing. Yet, these are only some of the advantages that computational tools in education offer: other areas of great potential involve networked user communities. Mechanisms to intelligently interconnect learners and educational communities are a valuable goal, and can significantly contribute to advanced knowledge discovery and sharing. A reasonable starting point for this is to use the artifacts created by the users/learners, and to derive potential interaction partners based on this source of information. However, there are number of practical problems with this approach:

- The computational tools used within education are massively heterogeneous – even within content-level or educational domains. A common data format that could allow for generic access by analyzing algorithms is neither available nor realistic. Thus, the re-use of material across tools is not possible, and even the detection of interesting data is hard.
- Metadata that complies with accepted standards could solve at least the detection problem stated above. Yet, the process of manually indexing data is inconvenient and time consuming, so that people tend to avoid it, as there is usually little direct benefit [1].
- Mechanisms for determining potential interaction partners based on the content-level similarity of heterogeneous data are rare and not easy to define, especially under the criterion of simplicity in usage – which is important in educational settings.

The initial idea for this paper relies on the concept of *communication through artifacts*, a principle originally rooted in shared workspaces scenarios [2]. One problem with the concept is that, in its original form, it is restricted to synchronous cooperation with few participants in shared workspace scenarios. This paper presents an approach to retain the effects of artifacts used for mediating collaboration, while relaxing the constraints of time and group size. The approach relies on the following key ideas: (1) an exploitation of task context, which is provided by tools and archive systems, (2) the support of heterogeneous applications and data types through indirect mechanisms that allow the connection of users even though their preferred document formats might be incompatible, and (3) a mix of retrieval mechanisms which integrates, e.g., recommendation mechanisms and ontology based querying.

2. Existing approaches and technologies

A frequently chosen representation technique to address semantic interoperability concerns between data formats are *ontologies*. These can be understood as conceptualizations of a domain into a human-understandable, but machine-readable format consisting of entities, attributes, relationships, and axioms [3]. This explains why metadata systems often make use of ontologies as underlying structures – compared to other less structured and formal approaches, their suitability for AI and knowledge representation techniques can significantly increase or supplement classical information retrieval techniques.
Research on *recommender systems*, which aim at proposing relevant documents to users, is probably the most closely related area to the approach presented in this paper. Yet, their foundations differ from the approach presented in this paper in that recommender systems typically rely on (user provided) *ratings* of documents, which are either used directly for recommendation of the rated document, or indirectly to infer ratings for similar documents. A frequently applied method in the field of recommender systems is *collaborative filtering*, which has proven to be effective. Yet, it has several inherent problems, including the cold start problem. Several mechanisms to overcome this problem have been proposed. Some base on the idea of community membership, while others [4] investigate the synergies evolving from an integration of recommender systems with ontologies. Recently, also an approach which uses document assessments to dynamically update user profiles and build communities based on these profiles has been proposed [5].

3. Approach and System Architecture

The driving idea for the approach proposed in this paper differs from the listed concepts in that it does not require neither explicit nor implicit document assessment, but instead makes use of automatically available activity contexts for indexing and retrieval. A first feasibility study [6] illustrated how functionality required can even be embedded in the tool that provides the task context, thereby linking of user- and document-related metadata (e.g., including educational or domain specific dimensions, tool information, and user roles).
To flexibly interconnect heterogeneous tools with archives, a layered architecture can serve as a technical backbone for implementing the ideas presented in this paper. The facilities for the connection of users are then typically contained in a medium layer, consisting of four core components: (1) a *bridge* between the tools and the archives to allow for document uploads and downloads, as well as for transmitting queries and recommendations, (2) an *ontology* to store relevant domain/educational concepts and interrelations, (3) a representation of *task context* as extracted from archives, tools, and the ontology, and (4) an *AI based Community support engine* (CSE) that is in charge of calculating recommendations for potential interaction partners. The basic interface function of the CSE has the following signature:

`recommend(doc)→[user_list]`

The algorithm analyses the input document (content & context) and calculates a list of potentially interested users. `recommend` uses two functions to calculate the desired output:

- `related(doc)→[doc_list]` determines a set of documents that are similar to the parameter document, while
- `profile(doc,user)→[rating_list]` evaluates one document against all the documents created by a given user.

Both `profile` and `related` rely on a basic function which estimates the similarity of documents (including contexts):

$\texttt{rate(doc}_1\texttt{,doc}_2\texttt{)}\rightarrow\texttt{rating}$

This `rate` function builds the core of the CSE. It calculates a distance measure between documents with associated task contexts. However, in a realistic heterogeneous usage of the system concerning tools, data types, and probably a large number of missing or erroneous metadata pieces, good prediction qualities for the recommendation function are not easy to obtain, which motivates a more solid multi-method mix for the implementation of `rate`. This is addressed with the CSE structure illustrated in Figure 1.

- Even if the participating persons are new to the system and thus neither have a profile nor a document history, and the content data types are unknown to the system, the *ontology-driven engine* can estimate thematic proximity (in domain and educational terms) of data based on semantic context information.
- The *user profile comparison* takes general user/learner profiles (if available) as parameters, and allows for taking into account roles (e.g., teacher vs. student) or the languages of users. If available, also specific educational parameters of the user model can be taken into account here.
- Tool compatibility is an important information within the system, both directly (tool similarity calculated based on typology of tools), and also indirectly (data formats based, including compatibility information), as it allows for direct re-use of documents – this is addressed by the *tool-based similarity calculation*.
- The *archive distance measurement* uses inference techniques on an internal document/user graph to reveal document similarity of the type "most users who have seen document A have also looked at document B".
- The *inner similarity check* allows tools to define content-level distance measures, and thus to incorporate domain-specific knowledge in the calculation process.

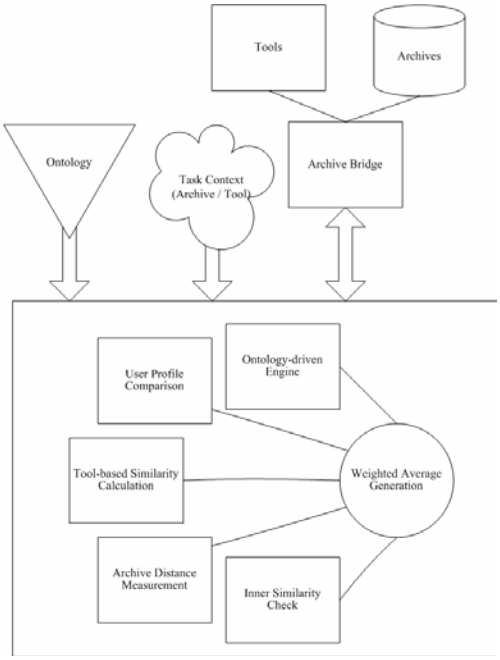

Figure 1. Architecture of the Community Support Engine

Finally, the *weighted average component* calculates a ranked user list based on the other single results – this list can then be used to recommend potential interaction partners.

References

[1] Wickens, C. D. (1992). *Engineering Psychology and Human Performance*. New York (NY), USA: Harper Collins.

[2] Dix, A.., Finlay, J., Abowd, G. D., & Beale, R. (2004). *Human-computer interaction*. Harlow, England: Pearson Education Limited.

[3] Guarino, N., & Giaretta, P. (1995). Ontologies and Knowledge Bases: towards a terminological clarification. In N. J. I. Mars (Ed.), *Towards very large knowledge bases: knowledge building and knowledge sharing*. Amsterdam: IOS Press.

[4] Middleton, S. E., Alani, H., & De Roure, D. C. (2002). Exploiting Synergy between Ontologies and Recommender Systems. In *Workshop Proceedings of WWW 2002*. Honolulu (HI), USA.

[5] Francq, P., & Delchambre, A. (2005). Using Document Assessment to Build Communities of Interest. In *Proceedings of SAINT 2005* (p. 327-333). IEEE Press.

[6] Pinkwart, N., Jansen, M., Oelinger, M., Korchounova, L., & Hoppe, H. U. (2004). Partial generation of contextualized metadata in a collaborative modeling environment. In L. Aroyo and C. Tasso (Eds.), *Workshop proceedings of AH 2004* (p. 372-376). Eindhoven (NL): Technical University Eindhoven.