Content Recommendation in APOSDLE using the Associative Network

Hermann Stern
(Graz University of Technology, Graz, Austria
hstern@tugraz.at)

Günter Beham
(Know-Center GmbH, Austria
gbeham@know-center.at)

Peter Kraker
(Know-Center GmbH, Austria
pkraker@know-center.at)

Stefanie N. Lindstaedt
(Know-Center GmbH, Austria
slind@know-center.at)

1 Extended Abstract

Work Integrated Learning (Lindstaedt and Farmer, 2004) (Smith, 2003) enables knowledge workers to learn directly at their place of work, by doing their usual tasks instead of exercising dedicated artificial learning situations, with working and learning seamlessly integrated (Lindstaedt et al., 2009). One of the key factors of Work Integrated Learning Environments is to provide the right content to the learner, both suitable for his current level of knowledge, and the task he is currently working on.

To be able to recommend relevant resources, the system has to know about (i) the knowledge worker himself, for example his experience in certain topics, and (ii) the current activity he is involved in – both together defining the working context. The work presented here originates in the EU project APOSDLE[1], which focuses on Work Integrated Learning. The detailed approach of the whole project is described in (Lindstaedt et al., 2007).

Content recommendation using the Associative Network is a core component of the APOSDLE retrieval system. It is responsible for providing content resources to the "Suggest" functionality, which recommends all types of resources (like documents, users etc.) to help the user with learning and executing his current task.

[1] Advanced Process Oriented and Self-Directed Learning Environment (http://www.aposdle.org)

APOSDLE Meta Model

APOSDLE uses a meta model individually created for the respective application domain which contains:

i. a task model (a formal description of work tasks to be supported, like "how to create a workshop");
ii. a domain model (a semantic description of the domain in terms of concepts, relations, and objects); and
iii. a learning goal model (providing a mapping between domain concepts, tasks and general learning goal types) (Lindstaedt et al., 2008).

Entities present in this meta model are used to annotate documents stored in the APOSDLE system to make them available for retrieval. Certain parts of these documents are tagged with concepts (the topic the selected part of the document is dealing with) and material uses (describing the type of the selected part, like introduction, bullet list, how-to etc.) yielding a list of annotated text fragments. In APOSDLE we call them "snippets".
Snippets can be created manually by the users themselves. Since this is a rather time consuming task, there is also the possibility of automatic assignment of this semantic metadata based on a text classification approach. It does, however, not equal the precision of metadata assigned by a human domain expert yet.

Associative Network

The Associative Network is the component in APOSDLE responsible for maintaining and retrieving snippets; in this network every concept from the domain model and every snippet in the system are represented by nodes. Three directly dependent layers provide the connection between the input (one or more topics) - and the list of snippets provided as output (see Figure 1).

The first layer (concept to concept layer) contains only domain concept nodes which are connected with edges by means of their semantic similarity; edge weights state the grade of similarity. This layer can be used to extend the list of input topics to also take other topics into account which are very similar but have not been selected by the user. The second layer (concept to document layer) states the mappings between the domain concepts and the snippets, annotating each snippet with exactly one topic. This is the only mandatory layer in the Associative Network since this is the core layer serving snippets for one or more topics; it cannot be turned off optionally. The third layer (document to document layer) contains only snippets and connects them with edges based on their textual similarity. If activated, this layer is adding similar snippets to the set of snippets coming from the second layer.

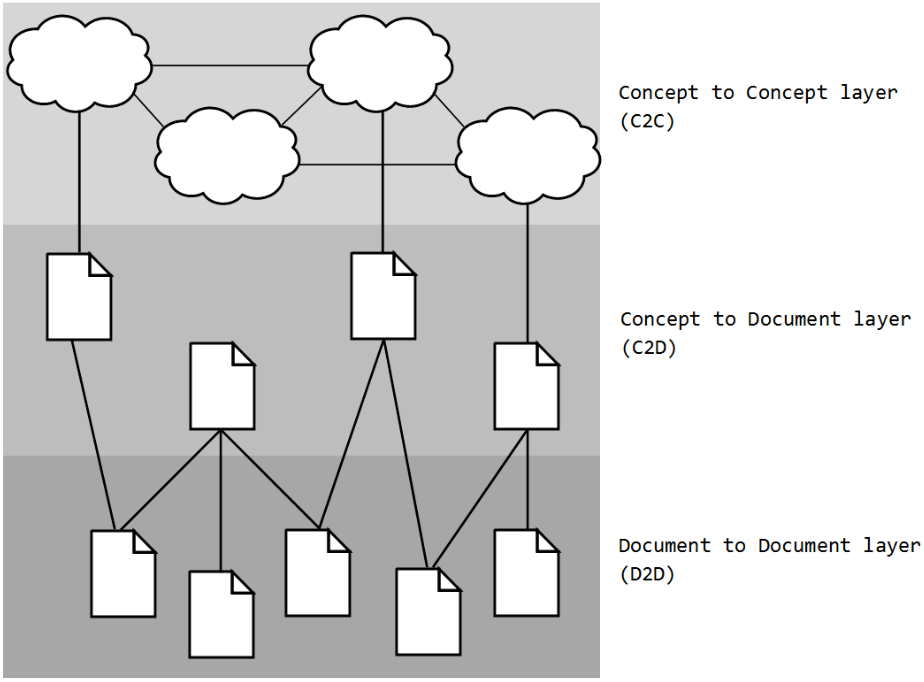

Figure 1 - Anatomy of the Associative Network

For processing the information in the network representation of layer one and three, a processing technique called Spreading Activation is used. Starting from a set of initially activated nodes in the net, the activation spreads over the network (Sharifian and Samani, 1997). During search, energy flows from a set of initially activated information items over the edges to their neighbours. (Lindstaedt et al., 2008)

The idea behind the Associative Network is to return a list of most relevant snippets to the user; the input concept (like "Workshop") (i) can be chosen by the user, (ii) is also detected automatically by the APOSDLE context detection functionality, or (iii) is inferred by a working task the user chooses. The Associative Network then takes this starting concept, uses the first layer (if turned on) to augment the list of starting concepts, gets all documents associated to these topics from layer two and finally augments the list of resulting documents in layer three (if turned on).

Compared to a traditional retrieval engine, the Associative Network introduces several advantages when dealing with sparse annotations. Since manual annotation of snippets is very time consuming and has to be done by domain experts, in most case only few concepts from the models are used for annotation, and only few resources are really annotated. The Associative Network is able to retrieve documents from concepts not directly selected by the user, but similar to them (concept to concept layer), and also documents that are not tagged with the concept, but similar to the found documents (document to document layer).

Evaluation

As the APOSDLE project now comes to its end, with the final prototype being implemented, we are in the process of evaluating several components. The Associative Network in APOSDLE will be evaluated concerning two major aspects:

First of all the semantic similarity measures used to weigh the edges in the concept to concept layer will be evaluated. The Associative Network uses this layer by taking the original input concept(s) and adding the most relevant concepts in the neighbourhood to the list of input values. Only one expansion step is carried out, so only directly connected neighbours are affected.
APOSDLE can exploit several measures to describe the similarity between two concepts of the domain ontology, in particular (i) Conceptual Similarity, (ii) Shortest Path, (iii) Scaled Shortest Path, (iv) Resnik, (v) Lin, (vi) Jian Conrath and (vii) Trento Vector.

Since there is no optimal measure for all of the domains available yet, it will have to be derived individually for each application domain. Representative for all the APOSDLE domains, two application domains (called ISN and CCI) will be evaluated in a lab study with the corresponding application partners. To keep place and time open to the involved domain experts, a Web application called Web Tool (for Ontology) Evaluation (WTE) will be provided.

In WTE a single rating session will consist of a fixed number of iterations. Random pairs of two concepts will be presented and the domain expert will have to rate the degree of interconnection between the two concepts. Using these ratings, an ideal model of the concept to concept layer will be created.
To ensure a correct understanding of the concepts shown, the whole path of the concepts up to the domain root will be visualized.

Table 1 shows the rating scale we are planning to use according to (Charles, 2000). To ensure consistent rating of all the participants, both the numbers and the description will be visible during the rating process.

Scale	*Description*
4	identical in meaning
3	similar in meaning
2	vaguely similar in meaning
1	different in meaning
0	opposite in meaning

Table 1 - Rating scale for concept pairs evaluation

It will be impossible to evaluate all the available edges - every concept in the domain ontology could possible have an edge to every other concept. The number of edges which can be really evaluated heavily depends on the manpower available for evaluation. Furthermore a random selection of concept pairs for rating does no take into account that some of them might be more significant for the domain than others.
We foresee doing some manual pre selection of approximately 20 percent of the concept pairs to be rated in one session, the other 80 percent will be selected randomly.
The aim of the pre selection will be to include concept pairs which will be useful to discriminate between measures; pairs that score very different values for the different measures will help us in finding the measure most similar to the human rating process.

We foresee two experts for every domain working with the Web application, One session will contain 100 concept pairs to be rated, taking about 15 to 20 minutes. The number of individual sessions will be up to the domain expert, the more the better of course.

Afterwards, the results of the ratings will be evaluated against the computed measures and the following questions will be answered:

i. Which measure one is best for a specific application domain?

ii. Is there a specific similarity measure which performs well on every domain tested and could therefore be proposed as default measure?

iii. How much do the measures and the results actually differ?

The second evaluation will deal with effectiveness of the classifier used to automatically tag snippets with topics. In this context, effectiveness is defined as the ability to take the right classification decisions (Sebastiani, 2002). A cross-validation on the existing manual annotations will be conducted. Several test runs with varying classifier parameters and different sets of concepts will be performed. For every test run, established measures of text categorization effectiveness like precision and recall will be computed and stored.

These measures will be used to answer three questions:

i. What is the categorization effectiveness of the concept classification used in the automatic annotation process?

ii. How does categorization effectiveness differ with varying classifier parameters given a certain distribution of manual annotations?

iii. What impact does the degree of manual annotation have on categorization effectiveness?

The Associative Network approach in APOSDLE will be used in other projects like MATURE[2], with several improvements planned: the network will be for example be able to store not only snippets, but all types of resources. The order of topics submitted as input will be taken into account; additional weighting algorithms and similarity measures will be developed.

Acknowledgements

APOSDLE is partially funded under the FP6 of the European Commission within the IST Workprogramme (project number 027023). The Know-Center is funded within the Austrian COMET Program - Competence Centers for Excellent Technologies - under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry of Economy, Family and Youth and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency FFG.

[2] http://mature-ip.eu/

References

[Anderson, 1983] Anderson, J. R.: A spreading activation theory of memory. Journal of Verbal Leaning and Verbal Behaviour, *22*, 1983, pp 261-295.

[Charles, 2000] Charles, W.G.: Contextual correlates of meaning, Applied Psycholinguistics, v21 n4 p505-24, Dec 2000.

[Lindstaedt and Farmer, 2004] Lindstaedt, S.N. & Farmer, J.: Kooperatives Lernen in Organisationen, in CSCLKompendium - Lehr- und Handbuch für das computerunterstützte kooperative Lernen, Oldenbourg Wissenschaftsverlag, München, Germany.

[Lindstaedt et al., 2007] Lindstaedt, S. N., Ley, T., Mayer, H.: APOSDLE - New Ways to Work, Learn and Collaborate, in Proceedings of the 4th Conference on Professional Knowledge Management WM2007, 28. - 30. März 2007, Potsdam, Germany, pp 381-382, GITO-Verlag, Berlin.

[Lindstaedt et al., 2008] Lindstaedt, S. N., Scheir, P., Lokaiczyk, R., Kump, B., Beham, G., Pammer, V.: Knowledge Services for Work-integrated Learning**,** Proceedings of the European Conference on Technology Enhanced Learning (ECTEL) 2008, Maastricht, The Netherlands, September 16-19, pp 234-244.

[Lindstaedt et al., 2009] Lindstaedt, S. N., Beham, G., Kump, B., Ley, T.: Getting to Know Your User – Unobtrusive User Model Maintenance within Work-Integrated Learning Environments, accepted for the EC-TEL 2009.

[Sebastiani, 2002] Sebastiani, F.: Machine Learning in Automated Text Categorization. ACM Computing Surveys, 34 (1), 2002, pp 1-47.

[Sharifian and Samani, 1997] Sharifian, F. & Samani, R.: Hierarchical spreading of activation, in Farzad Sharifian, ed., Proc. of the Conference on Language, Cognition, and Interpretation, IAU Press, 1997, pp 1-10.

[Smith, 2003] Smith, P. J.: Workplace Learning and Flexible Delivery. In: Review of Educational Research, 73, 1 (2003) pp 53-88.