

Knowledge map creation and maintenance for virtual communities of practice

Fu-ren Lin^a, and Chih-ming Hsueh^b

^aInstitute of Technology Management, National Tsing-hua University, Hsinchu 300, Taiwan, ROC

^bDepartment of Information Management, National Sun Yat-sen University, Kaohsiung 804, Taiwan, ROC

Received 13 May 2003; accepted 21 March 2005. Available online 8 June 2005.

Abstract

This paper proposes a knowledge map management system to facilitate **knowledge management** in virtual communities of practice. To realize the proposed knowledge map management, we develop knowledge map creation and maintenance functions by utilizing information retrieval and data mining techniques. The knowledge maps created respectively from the documents of the teachers' professional community, SCTNet, and the thesis repository at Taiwan's National Central **Library**, are evaluated by experts of these two domains. Knowledge maps generated by the system are accepted by domain experts from the evaluation since the degree of their modification of the automatically created knowledge maps is proportionally small. The knowledge structure representing the categories of community documents maintains its high purity, diversity, specificity, and structure adaptation by using the knowledge map maintenance function with limited computational cost. Thus, the knowledge map creation and maintenance mechanisms developed in this research enable the dynamic **knowledge management** of communities of practice on the Internet.

Keywords: Knowledge map; Information retrieval; Clustering; Community of practice; Professional community

Article Outline

1. [Introduction](#)
2. [Existing techniques used for knowledge map management](#)
 - 2.1. [Information retrieval](#)
 - 2.2. [Document clustering](#)
3. [The architecture of knowledge map management system](#)
 - 3.1. [Knowledge navigation](#)
 - 3.2. [Knowledge seeker](#)
 - 3.3. [Learning adviser](#)
 - 3.4. [Knowledge map manager](#)
4. [Knowledge map creation and maintenance](#)
 - 4.1. [Knowledge map creation](#)
 - 4.1.1. [Stage 1 \(hierarchical clustering\)](#)

- 4.1.2. [Stage 2 \(*k*-means clustering\)](#)
- 4.2. [Knowledge map maintenance](#)
 - 4.2.1. [Document and cluster representation](#)
 - 4.2.2. [Document classification](#)
 - 4.2.3. [Cluster decomposition](#)
 - 4.2.4. [Cluster merging](#)
 - 4.2.5. [Re-clustering](#)
- 5. [Evaluation of the knowledge map creation](#)
 - 5.1. [Evaluation criteria](#)
 - 5.2. [Experimental design](#)
 - 5.3. [Evaluation results](#)
- 6. [Evaluation of the knowledge map maintenance](#)
 - 6.1. [Evaluation criteria](#)
 - 6.2. [Evaluation results](#)
- 7. [Conclusions and future research](#)
- [References](#)

Table 1.
Interpretation of Silhouette Coefficient ([Kaufman and Rousseeuw, 1990](#))

Silhouette Coefficient	Interpretation
0.71–1.00	A strong structure has been found
0.51–0.70	A reasonable structure has been found
0.26–0.50	The structure is weak and could be artificial; please try additional methods on this data set
0.26	No substantial structure has been found

Table 2.
Precision/recall of NSD document set

	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	Expert 8	Average
Wide	0.90/0.91	0.85/0.85	0.91/0.91	0.90/0.92	0.93/0.93	0.95/0.94	0.92/0.92	0.92/0.92	0.91/0.91
Deep	0.86/0.88	0.91/0.88	0.91/0.93	0.97/0.97	0.96/0.97	0.98/0.98	0.91/0.92	0.92/0.91	0.93/0.93
Mix	0.95/0.95	0.91/0.88	0.89/0.91	0.98/0.98	0.98/0.98	1.00/1.00	0.91/0.91	0.93/0.91	0.95/0.94

Table 3.
Precision/recall of TAIM document set

	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	Expert 8	Average
Wide	0.94/0.94	0.95/0.95	0.94/0.95	0.95/0.95	0.96/0.96	0.94/0.94	0.89/0.88	0.91/0.91	0.93/0.93
Deep	0.92/0.92	0.93/0.93	0.94/0.94	0.93/0.93	0.97/0.97	0.88/0.97	0.87/0.87	0.87/0.87	0.91/0.91

	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	Expert 8	Average
	2	3	4	4	6	1	7	8	2
Mix	0.95/0.9 6	0.97/0.9 7	0.90/0.9 1	0.98/0.9 8	0.96/0.9 4	0.90/0.8 9	0.82/0.8 2	0.92/0.9 1	0.93/0.9 2

Table 4.

Evaluation results of incremental knowledge map maintenance for NSD and TAIM document sets

	Purity	Diversity	Specificity	SAI	# of re-clustering
NSD ($\alpha = 0.84, \beta = -10\%$)	65%	74%	81%	55	4/85 (16/254) ^a
TAIM ($\alpha = 0.84, \beta = -10\%$)	75%	72%	88%	39	2/94 (13/281)

^a 4/85(16/254) means that four times of re-clustering occur out of 85 document insertions, and 16 out of 254 documents on average are re-categorized.

References

- Agrawal et al., 1999 Agrawal, R., Bayardo, R., & Srikant, R. (1999). Athena: Mining-based interactive management of text databases. In *Proceedings of the seventh conference on extending database technology, July*.
- Agrawal et al., 1998 Agrawal, R., Gehrke, J., Gunopulos, D., & Raghavan, P. (1998). Automatic subspace clustering of high dimensional data for data mining applications. In *SIGMOD'98*.
- Baeza-Yates and Gonnet, 1996 R. Baeza-Yates and G. Gonnet, Fast text searching for regular expressions or automaton searching on tries, *Journal of the ACM* **43** (1996) (6), pp. 915–936. [Abstract-Compendex](#) | [Abstract-INSPEC](#) | [\\$Order Document](#) | [MathSciNet](#) | [Full Text via CrossRef](#)
- Biru et al., 1989 T. Biru, A. El-Hamdouchi, R.S. Rees and P. Willett, Inclusion of relevance information in the term discrimination model, *Journal of Documentation* **45** (1989), pp. 85–109. [Abstract-INSPEC](#) | [\\$Order Document](#)
- Browne et al., 1997 G. Browne, S. Curley and P. Benson, Evoking information in probability assessment: Knowledge maps and reasoning-based directed questions, *Managements Science* **43** (1997) (1), pp. 1–14. [Abstract-INSPEC](#) | [\\$Order Document](#)
- Chien, 1997 Chien, L.-f. (1997). PAT-tree-based keyword extraction for Chinese information retrieval. In *Proceedings of the 1997 ACM SIGIR, Philadelphia, PA, USA* (pp. 50–58).
- Chien and Pu, 1996 L.-f. Chien and H.-t. Pu, Important issues on Chinese information retrieval, *Computational Linguistics and Chinese Language Processing* **1** (1996) (1), pp. 205–221.

Chou and Lin, 1998 C. Chou and H. Lin, The effects of navigation map types and cognitive styles on learners' performance in a computer-networked hypertext learning system, *Journal of Educational Multimedia and Hypermedia* 7 (1998), pp. 151–176.

Gonnet and Baeza-Yates, 1992 G.H. Gonnet and R. Baeza-Yates, New indices for text: PAT trees and PAT arrays, *Information retrieval data structures and algorithms*, Prentice Hall (1992), pp. 66–82.

Kaufman and Rousseeuw, 1990 L. Kaufman and P.J. Rousseeuw, Finding group in data: An introduction to cluster analysis, John Wiley and Sons, Inc (1990).

Knuth, 1973 D.E. Knuth, *The art of computer programming: Sorting and searching vol. 3*, Addison-Wesley, MA (1973).

Kohonen et al., 2000 T. Kohonen, S. Kaski, K. Lagus, J. Salojvi, V. Paatero and A. Saarela, Self organization of a massive document collection, *IEEE Transactions on Neural Networks* 11 (2000) (3), pp. 574–585. [Abstract-Compendex](#) | [Abstract-INSPEC](#) | [\\$Order Document](#) | [Full Text via CrossRef](#)

Li and Xing, 1998 Li, Z., & Xing, L. (1998). Search the Chinese Web-design and the operation of net-compass. In *Proceedings of the first Asia digital library workshop* (pp. 42–46).

Lin and Lin, 2001 F.-r. Lin and S.-c. Lin, A conceptual model for virtual organizational learning, *Journal of Organizational Computing and Electronic Commerce* 11 (2001) (3), pp. 155–178. [Abstract-PsycINFO](#) | [\\$Order Document](#)

Merkel and Rauber, 1999 Merkl, D., & Rauber, A. (1999). Automatic labeling of self-organizing maps for information retrieval. In *Proceedings of ICONIP '99, 6th international conference* (pp. 37–42).

Morrison, 1968 D. Morrison, PATRICIA: Practical algorithm to retrieve information coded in alphanumeric, *Journal of ACM*(1968), pp. 514–534. [Full Text via CrossRef](#)

Nonaka, 1994 I. Nonaka, A dynamic theory of organizational knowledge creation, *Organization Science* 5 (1994) (1), pp. 14–37. [Abstract-PsycINFO](#) | [\\$Order Document](#)

Paolucci, 1998 R. Paolucci, The effects of cognitive style and knowledge structure on performance using a hypermedia learning system, *Journal of Educational Multimedia and Hypermedia* 7 (1998), pp. 123–150. [Abstract-INSPEC](#) | [\\$Order Document](#)

Pask and Scott, 1972 G. Pask and B.C.E. Scott, Learning strategies and individual competence, *International Journal of Man–Machine Studies* 4 (1972), pp. 217–253. [Abstract-Compendex](#) | [Abstract-PsycINFO](#) | [\\$Order Document](#)

Punj and Stewart, 1983 G. Punj and D. Stewart, Cluster analysis in marketing research: Review and suggestions for application, *Journal of Marketing Research* (1983) (May), pp. 134–148. [Abstract-PsycINFO](#) | [\\$Order Document](#)

Roussinov and Chen, 1999 D. Roussinov and H. Chen, Document clustering for electronic meetings: an experimental comparison of two techniques, *Decision Support Systems* **27** (1999) (1–2), pp. 67–80.

Salton, 1989 G. Salton, Automatic text processing, Addison-Wesley, Reading, MA (1989).

Salton and Buckley, 1975 G. Salton and C. Buckley, A vector-space model for automatic indexing, *Communications of the ACM* **18** (1975) (11), pp. 613–620. [Abstract-Compendex](#) | [Abstract-INSPEC](#) | [\\$Order Document](#) | [Full Text via CrossRef](#)

Salton and Buckley, 1988 G. Salton and C. Buckley, Term-weighting approaches in automatic text retrieval, *Information Processing and Management* **24** (1988) (5), pp. 513–523. [Abstract](#) | [Abstract + References](#) | [PDF \(1050 K\)](#)

Sebastiani, 2002 F. Sebastiani, Machine learning in automated text categorization, *ACM Computing Survey* **34** (2002) (1), pp. 1–47. [Abstract-Compendex](#) | [Abstract-INSPEC](#) | [\\$Order Document](#) | [Full Text via CrossRef](#)

Steinbach et al., 2000 Steinbach, M., Karypis, G., & Kumar, V. (2000). A comparison of document clustering techniques. *KDD-2000 workshop on text mining, August 20, Boston, MA*. Available from http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&locato r=url&cdi=5948&plusSign=%2B&targetURL=http%253A%252F%252Fcite seer.nj.nec.com%252Fsteinbach00comparison.html.

Wei et al., 2002 C.-p. Wei, P.J. Hu and Y.-x. Dong, Managing document categories in e-commerce environments: An evolution-based approach, *European Journal of Information Systems* **11** (2002) (3), pp. 208–222. [Abstract-INSPEC](#) | [\\$Order Document](#)

Witkin et al., 1977 H.A. Witkin, C.A. Moore, D.R. Goodenough and P.W. Cox, Field-dependent and field-independent cognitive styles and their educational implications, *Review of Educational Research* **47** (1977), pp. 1–64. [Abstract-PsycINFO](#) | [\\$Order Document](#)

Wong and Li, 1998 Wong, K.-f., & Li, W.-j. (1998). Intelligent Chinese information retrieval—why is it so difficult? In *Proceedings of the first Asia digital library workshop*.

Wu and Tseng, 1993 Z. Wu and G. Tseng, Chinese text segmentation for text retrieval: Achievements and problems, *Journal of the American Society for Information Sciences* **44** (1993), pp. 532–542. [Abstract-INSPEC](#) | [\\$Order Document](#) | [Full Text via CrossRef](#)

Wu and Tseng, 1995 Z. Wu and G. Tseng, ACTS: an automatic Chinese text segmentation system for full text retrieval, *Journal of the American Society for Information Sciences* **46** (1995) (2), pp. 532–542.

Yang and Lee, 1999 Yang, H., & Lee, C. (1999). A text data mining approach using a Chinese corpus based on self-organizing map. In *The fourth international workshop on information retrieval with Asian languages*.

Yang et al., 1998 Yang, C., Yen, J., Yung, S., & Chung, A. (1998). Chinese indexing using mutual information. In *Proceedings of the first Asia digital library workshop* (pp. 57–64).