Weak Information Structure for Human Information Sharing

Harumi Maeda, Masao Kajihara, Hidekazu Adachi
Atsushi Sawada, Hideaki Takeda and **Toyoaki Nishida**
Graduate School of Information Science
Nara Institute of Science and Technology
8916-5, Takayama, Ikoma, Nara 630-01, Japan
Phone: +81-7437-2-5265 Fax: +81-7437-2-5269
E-mail: harumi-m@is.aist-nara.ac.jp

Abstract

This paper demonstrates the effectiveness of the *weak information structure* for *human information sharing.* The weak information structure is an information representation which connects various information media without defining the semantics rigorously. By leaving the interpretation of the semantics to tacit human background knowledge, it becomes compact and robust. To investigate how effective the weak information structure is in real world problems, we test two cases and report our results. (1) We have developed a system called CM-2 which gathers and reorganizes information on the Internet. Concerning precision and recall rate, the results are between 68–90%. (2) We have developed a system called InfoCommon which supports information sharing in community. More than 50% of users answer that the system is useful for getting information they need.

1 Introduction

World Wide Web has become popular for human information sharing on the Internet. As large-scale information resources on the Internet are increasing rapidly, it becomes more and more difficult to obtain information we need. Although a number of search tools are available, there are few intelligent systems which help us reorganize vast information obtained from the Internet. We point out that this results from conceptual diversity in HTML documents on WWW.

On the other hand, programming languages or knowledge representation such as first-order logic or frames have been used for computer information sharing. Unfortunately, these are so logically rigid that one must spend tremendous amount of efforts on information acquisition. This defect forces too much on human effort and hence significantly hinders accumulation of a large amount of useful information.

We investigate an intermediate information representation which is weaker than well-defined knowledge representation for human information sharing.

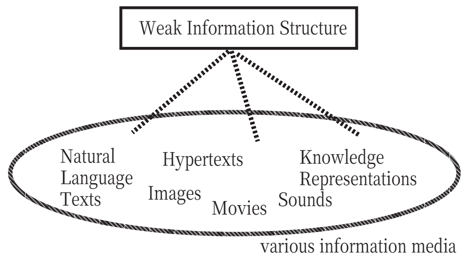


Figure 1: The Weak Information Structure

In this paper, we propose that the *weak information structure* is effective for *human information sharing.* The weak information structure is an information representation which connects a wide variety of information media such as natural language texts, hypertexts and images without defining the semantics rigorously (Figure 1). By leaving the interpretation of the semantics to tacit human background knowledge inherently shared with people, it becomes compact and robust. Moreover, the weak information structure is easy to generate from raw data for both of humans and computers, and therefore can reduce the cost of information acquisition and integration.

To investigate how effective the weak information structure is in real world problems, we test two cases and report our results: (1) gathering and reorganization of information on the Internet and (2) information sharing in community.

In what follows, first we describe two cases of experiments and then make discussion.

2 Test Case 1: Information Gathering and Reorganization on the Internet

It is difficult to find relevant information from large-scale information resources on the Internet. To integrate a wide variety of diverse information on the Internet, we developed a system called CM-2[1]: a system for information gathering and reorganization.

[1] "CM" stands for "Contextual Media" which stands for our long term theoretical research goal.

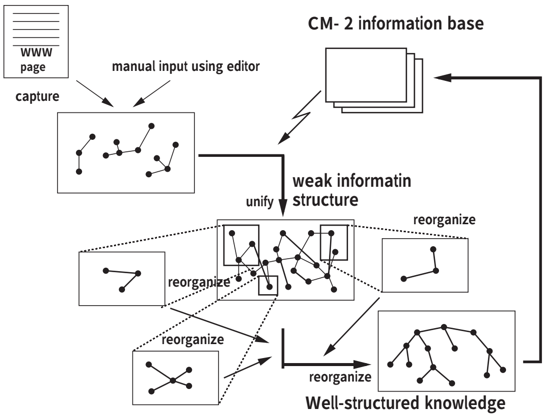


Figure 2: Overview of CM-2

2.1 Overview of CM-2

CM-2 consists of a collection of information bases. Each CM-2 information base is possessed by an individual person or a group and it consists of a collection of *workspaces* and *agents*. Each workspace provides a particular view of multimedia information stored in the information base. Each agent manipulates information tasks and interacts with the user. The user or the agents can interact with other, or incorporate information from other kinds of information sources connected to the Internet.

The basic entities of the weak information structure in CM-2 are a *unit* which represents either a concept or an external datum, and an *association* which connects a collection of key concepts with a collection of units which is normally reminded by the given keys.

CM-2 has following functions.

Information Gathering CM-2 generates units and associations from various information sources (e.g. HTML documents on WWW and newspaper databases) by using morphological analysis and heuristics and analyzing the structure of the documents.

Information Unification CM-2 unifies various associations such as generated from WWW pages and those constructed by humans into new information bases.

Information Reorganization Unified information bases often contain redundant or inconsistent information. Following functions help humans to edit and reorganize information bases. (1) Focus: to hide units which are unrelated to selected units, (2) Neighbor Search: to display units which are related to selected units, (3) Path Finding: to display relations between selected units, (4) Unit Search: to display selected units.

Information Decomposition CM-2 decomposes large information bases into small pieces which are easy to handle.

2.2 Experiment

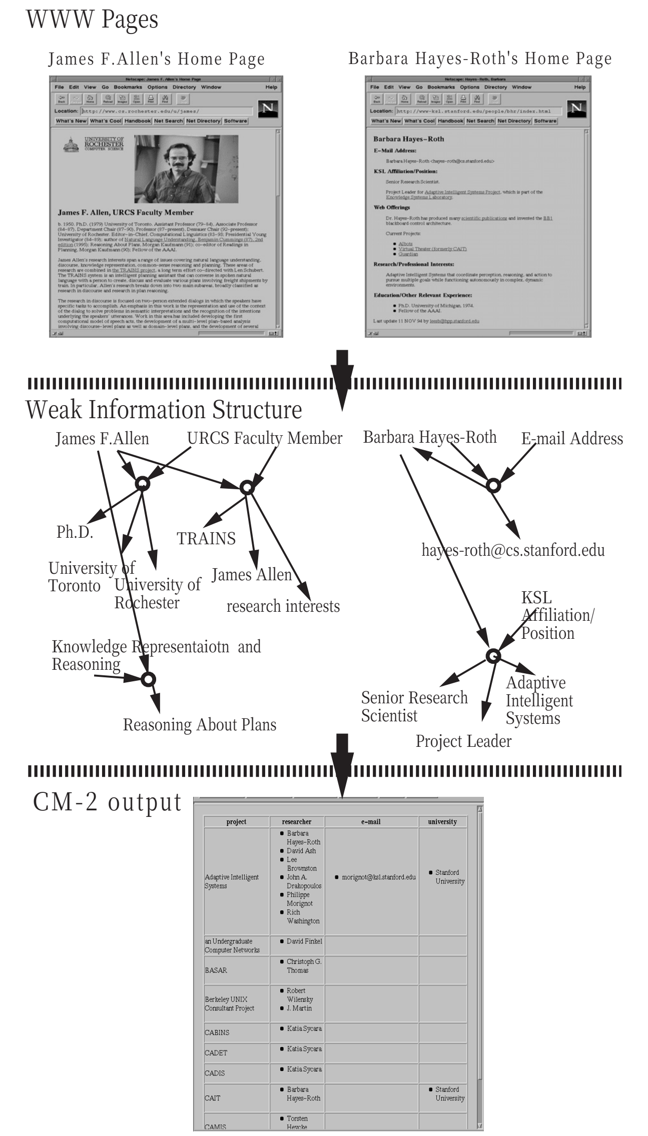


Figure 3: Overview of Information Gathering and Reorganization in CM-2

We gave 100 WWW pages concerning AI researchers to CM-2 for organizing AI directories. CM-2 extracted units about 7 classes (researchers, projects, e-mail, topics, universities, departments and laboratories), and generated associations. CM-2 reorganized these units and associations to display various lists according to user's input.

The overview of the process is illustrated in Figure 3. The algorithm is quite simple; (1) to generate units and associations by morphological analysis and analyzing HTML structure, (2) to identify class of the generated units using heuristics, (3) to unify units and associations using heuristics, and (4) to reorganize associations by path-finding according to user's input.

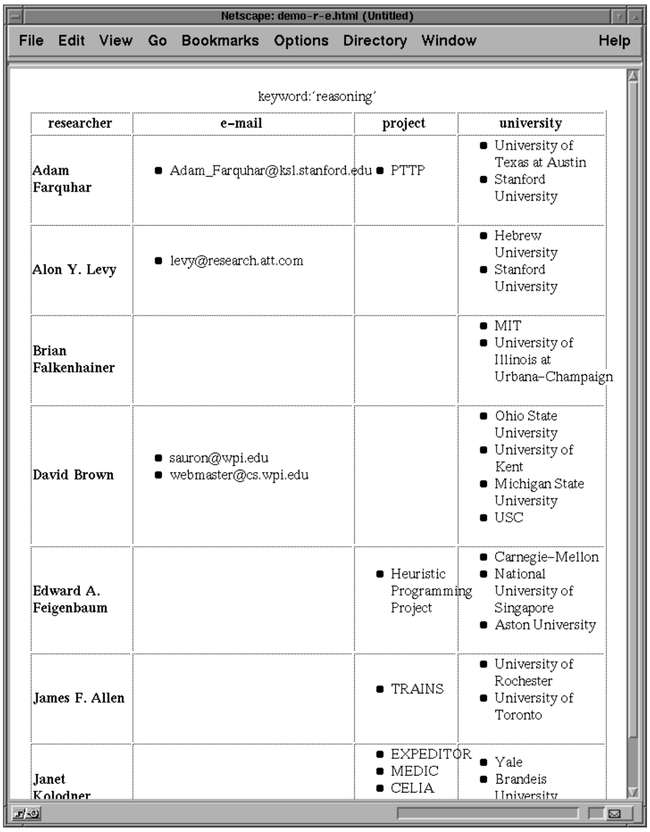


Figure 4: An Example Result of Information Reorganization in CM-2

Figure 4 shows an example result when a user input "reasoning","researcher","e-mail","project" and "university". For example, researchers such as Adam Farquhar, Alon Levy, Edward Feigenbaum and James Allen are extracted first, because the word "reasoning" and their names are written near in WWW pages, and then their related information is reorganized.

A summary of the results of two tests are shown in Table 1: (1) to display researchers' list (Test 1) and (2) to display projects' list (Test 2). The results of Test 1 (90% at precision rate, 83% at recall rate) are better than those of Test2 (68% at precision rate, 73% at recall rate), because original WWW pages are persons' pages.

From these results, we found that (1) the weak information structure is easy to generate from HTML documents and (2) reorganizing diverse information can be performed at reasonable rate for the rough algorithm by using the weak information structure and users' background knowledge.

2.3 Summary

To investigate how effective the weak information structure is for gathering and reorganizing information on the Internet, we developed a system called CM-2 which gathers and reorganizes information from WWW pages. We tested two cases: (1) to display researchers' list (Test 1) and (2) to display projects' list (Test 2). From this experiment, we found that (1) the weak information structure is easy to generate from HTML documents and (2) reorganizing diverse information can be performed at reasonable rate (68–90%) for the rough algorithm. This indicates that the use of the weak information structure and users' background knowledge is effective to extract and reorganize information from WWW pages.

Table 1: Result of Test Case 1

Test	Precision	Recall
Test 1 (researcher)	90%	83%
Test 2 (project)	68%	73%

Precision: $\frac{appropriate\ units}{generated\ units} \times 100$ (%)

Recall: $\frac{appropriate\ units}{units\ which\ should\ be\ extracted} \times 100$ (%)

3 Test Case 2: Information Sharing in Community

3.1 Overview of InfoCommon

Not only formal information, but also informal information such as word-of-mouth information is important to support information sharing in community. In order to integrate heterogeneous information for information sharing in community, we developed a shared card information system called InfoCommon. InfoCommon allows seamless keyword-based access to a variety of information cards to create a shared information base.

We employed several design principles to facilitate information sharing in community. First, we try not to enforce anyone particular concept. Instead, we allow much freedom in the usage of terms and the structure of shared information and to incorporate useful information from various viewpoints. Second, we make the information space a single seamless space. This releases the user from working with a rigid menu. Third, we enable the user to build a personal information space where he/she can organize relatively small amounts of information as desired[Nishida *et al.*, 1995].

InfoCommon provides visual interface for retrieving and sending information cards. Figure 6 shows a screen image of InfoCommon. The relation between two information cards is displayed by a link. InfoCommon information base consists of (a) a knowledge base which links keywords and information cards using the weak information structure and (b) an information card base as shown in Figure 7.

InfoCommon supports the following functions.

Content-based Information Retrieval Given a set of keywords, InfoCommon will respond with the set of information cards connected to the keywords. The result of retrieval is stored in the user's local information base where the user can re-arrange the collection of information cards, and add/remove nodes/links as desired.

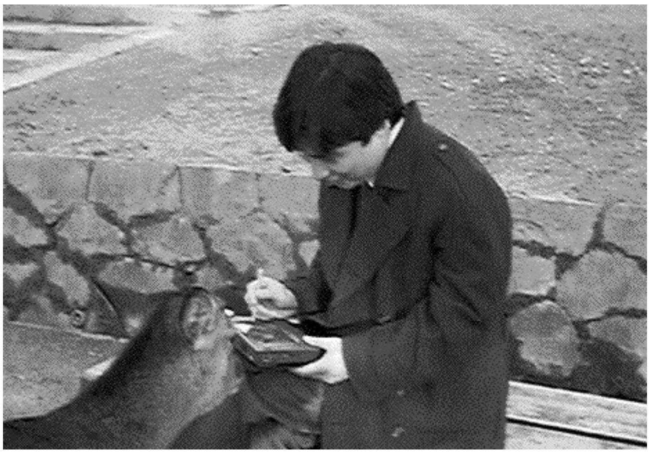


Figure 5: InfoCommon in Use

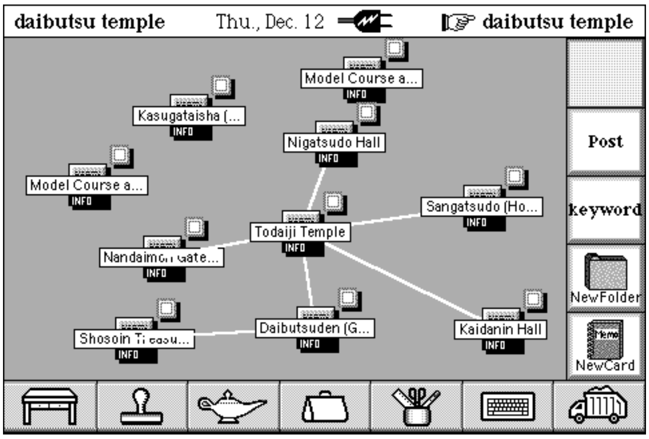


Figure 6: Screen Image of InfoCommon

Information Sending (Posting News) InfoCommon is built on a conventional News service. Users can add information cards on their local information base and send them to a News server.

3.2 Experiment

We evaluated the usefulness of InfoCommon at the ICMAS'96 (Second International Conference on Multiagent Systems) Mobile Assistant Project, which is the world first experiment in applying mobile computing systems to community support. 100 personal intelligent communicators with handy phones were loaned to conference participants to actually try out the system. Figure 5 is a photo taken in the Nara Park where an excursion was held, which shows how a user actually used InfoCommon.

InfoCommon information bases store static information such as abstracts of papers, session, local information and profiles to share information among participants.

The number of information retrievals was 351 and information sending was 32 in InfoCommon over the 5 day period.

We analyzed how InfoCommon was used by examing log files and questionnaires.

Purpose The answers to the question "For what did you used InfoCommon ?" is shown in Figure 8(a). 59% used the system for information retrieval and 19% for information sending including News, help desk and discussion. The reasons as to why InfoCommon was used for News were described as follows: "Because keyword search was easy and useful (14 persons)", "Because I found interesting topics in InfoCommon(5 persons)" and "Because I had a question (4 persons)". We determined that InfoCommon added new facility to conventional News reader. Major reasons of the choice "didn't use" were "slow information retrieval (8 persons)" and "I couldn't connect to the server (4 persons)". These problems involve server response and so are easily fixed.

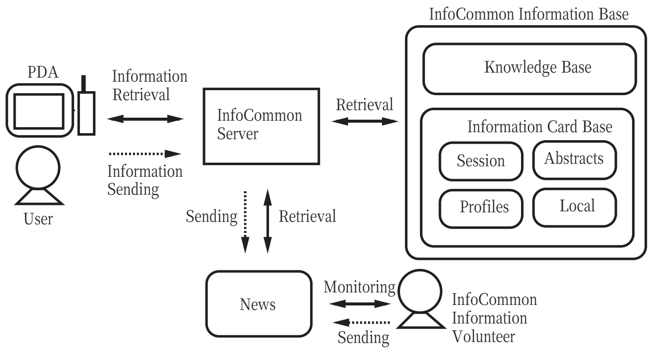
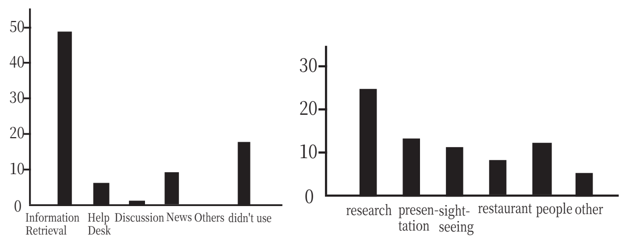


Figure 7: overview of InfoCommon

Topic Answers to the question "For what topic did you use InfoCommon" are displayed in Figure8(b).

Statistics The most frequent keyword input in information retrieval was "icstat"(Table2), a special keyword to display statistics concerning frequently asked keywords and frequently discussed subjects. This result suggests that users used InfoCommon to know what other people are interested in or what are hot before information retrieval or sending. We found that such statistics are useful for sharing information among participants.



(a)Purpose (b)Topic

Figure 8: Purpose and Topic

Table 2: Frequently Asked Keywords

ranking	keyword	ranking	keyword
1	icstat	7	lunch
2	nara	9	banquet
3	fipa	9	icsuggest
4	keihanna	9	kamameshi
5	agent	9	restaurant
6	nishimura	9	shuttle
7	food		

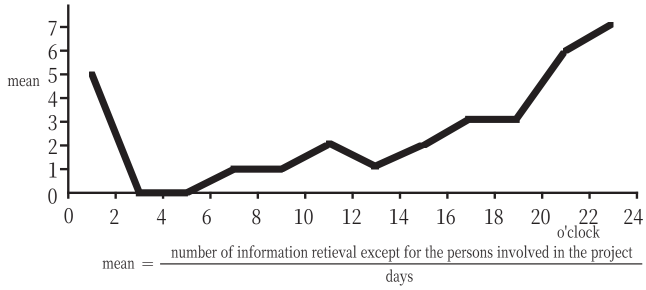


Figure 9: Change of Search

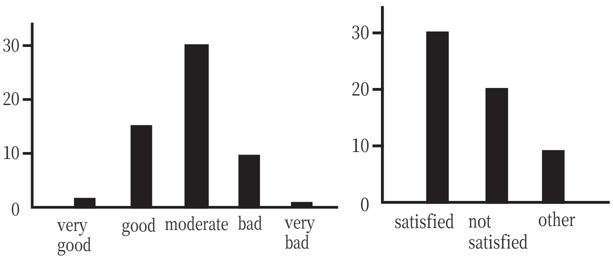
Change of Search Figure 9 shows that InfoCommon was used from evening to midnight. We analyze that the reasons are perhaps (1) people were busy during the daytime, (2) information retrieval was not a very urgent task, and (3) the system was slow.

Search Results 81% felt that the search results were fine(Figure10(a)).

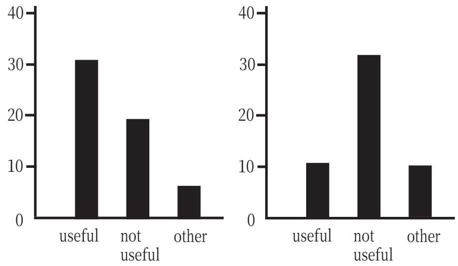
Satisfaction 51% answered that they were satisfied with InfoCommon(Figure 10(b)).

Usefulness 55% answered that InfoCommon was useful for getting the information they need. On the other hand, only 26% answered that it was useful for discussion.

We feel that the period (5 days) was too short to form the kind of community in which people create many active discussions. We need to conduct a longer-term experiment to evaluate the usefulness of the system for discussion in the future.



(a)Search Results (b)Satisfaction
Figure 10: Search Results and Satisfaction



(a)information retrieval (b)discussion
Figure 11: Usefulness

3.3 Summary

We developed a shard card information system called InfoCommon and evaluated it at the ICMAS'96 Mobile Assistant Project. The following results indicate the system's usefulness for information sharing in community: (a) 51% of users were satisfied of the system, (b) 81% felt search results were fine and (c) 55% answered that the system was useful for getting information they need. We consider the results are supporting evidence of the effectiveness of the weak information structure for information sharing in community.

4 Related Work and Discussion

CYC[Guha and Lenat, 1994] and ARPA Knowledge Sharing Effort[Patil *et al.*, 1992] have made a significant contribution in the sense they shed light on the importance of knowledge and information sharing and that they have presented a self-completed computational model. Their approach orients computer information sharing while our approach is for human information sharing.

Our work is related to recent work on information gathering from heterogeneous sources on the Internet ([Levy *et al.*, 1994],[Armstrong *et al.*, 1995],[Balabanovi'c and Shoham, 1995],[Li, 1995],[Iwazume *et al.*, 1996]). Instead of focusing on the strategies and heuristics for information gathering, we concentrate on how to classify information obtained from multiple information sources and integrate it into personal information base.

Gaines uses semantic networks as information representation for group knowledge sharing[Gaines and Shaw, 1994]. Our approach is based on much weaker information representation than semantic networks.

Kautz studied the use of agents in assisting and simplifying person-to-person communication for information gathering tasks[Kautz *et al.*, 1996]. They focus on the use of a software agent. We concentrate on the process of how humans create knowledge and information.

Sumi, Hori[Sumi *et al.*, 1992] and Kunifuji claim the importance of knowledge and information in the field of creative thinking support.

The basic recognition behind this research is a tradeoff between the benefit from conceptually well-structured information representation and the cost of information acquisition and integration. The more well-structured information representation becomes, the more useful it is for computational manipulation, however, the more expensive the cost of information acquisition and integration becomes for both of humans and computers.

Our approach is to provide a framework of collaborations for human information sharing with a low structural facilities.

The results of two test cases indicate that the weak information structure is effective for (1) information gathering and reorganization on the Internet and (2) information sharing in community.

We found that human background knowledge can be utilized in human information sharing. In addition, we also found that the weak information structure is easy to generate from raw data for both of humans and computers, and can reduce the cost of information acquisition and integration.

5 Conclusions

This paper demonstrated the effectiveness of the *weak information structure* for *human information sharing.* The weak information structure is an information representation which connects various information media without defining the semantics rigorously. By leaving the interpretation of the semantics to tacit human background knowledge, it becomes compact and robust. To investigate how effective the weak information structure is in real world problems, we tested two cases and reported our results. (1) We developed a system called CM-2 which gathers and reorganizes information on the Internet. Concerning precision and recall rate, the results were between 68–90%. (2) We developed a system called InfoCommon which supports information sharing in community. More than 50% of users answered that the system was useful for getting information they need.

As a future research, we plan to evaluate how people actually create the weak information structure and how people interpret the weak information structure created by others.

References

[Armstrong *et al.*, 1995] Robert Armstrong, , Dayne Freitag, Thorsten Joachims, and Tom Mitchell. WebWatcher: A learning apprentice for the World Wide Web. In *Working Notes of the AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments*, pages 6–12, 1995.

[Balabanovi'c and Shoham, 1995] Marko Balabanovi'c and Yoav Shoham. Learning information retrieval agents: Experiments with automated web browsing. In *Working Notes of the AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments*, pages 13–18, 1995.

[Gaines and Shaw, 1994] Brian R. Gaines and Mildred L. G. Shaw. Using knowledge acquisition and representation tools to support scientific communities. In *Proceedings AAAI-94*, pages 707–714, 1994.

[Guha and Lenat, 1994] R. V. Guha and Douglas B. Lenat. Enabling agents to work together. *Communications of ACM*, 37(7):127–142, 1994.

[Iwazume *et al.*, 1996] Michiaki Iwazume, Hideaki Takeda, and Toyoaki Nishida. Ontology-based information gathering and text categorization from the internet. In *Proceedings of the Ninth International Conference in Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*, 1996. 305–314.

[Kautz *et al.*, 1996] Henry Kautz, Bart Selman, and Al Milewski. Agent amplified communication. In *Proceedings of the Thirteen National Conference on Artificial Intelligence (AAAI-96)*, 1996. 3–9.

[Levy *et al.*, 1994] Alon Y. Levy, Yehoshua Sagiv, and Divesh Srivasava. Towards efficient information gathering agents. In *Working Notes of the AAAI Spring Symposium on Software Agents*, pages 64–70, 1994.

[Li, 1995] Wen-Syan Li. Knowledge gathering and matching in heterogeneous databases. In *Working Notes of the AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments*, pages 116–121, 1995.

[Nishida *et al.*, 1995] Toyoaki Nishida, Hideaki Takeda, Kenji Iino, and Masanobu Nishiki. A knowledge media approach to ontology development. In N.J.I. Mars, editor, *Towards Very Large Knowledge Bases: Knowledge Building & Knowledge Sharing 1995*, pages 84–94. IOS Press, 1995.

[Patil *et al.*, 1992] R. S. Patil, R. E. Fikes, P. F. Patel-Schneider, D. McKay, T. Finin, T. R. Gruber, and R. Neches. The DARPA knowledge sharing effort: Progress report. In Charles Rich, Bernhard Nebel, and William Swartout, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Third International Conference*. Morgan Kaufmann, 1992.

[Sumi *et al.*, 1992] Yasuyuki Sumi, Koichi Hori, and Setso Ohsuga. Computer-aided thinking based on mapping text-objects into metric spaces. In *Proc. of The 2nd Pacific Rim International Conference on Artificial Intelligence*, 1992.