A Plain Indexing Method for Organizing Conceptually Promiscuous Data

Toyoaki Nishida, Kazuto Koujitani, and Hideaki Takeda
Graduate School of Information Science
Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara 630-01, Japan
email: nishida@is.aist-nara.ac.jp

Abstract

An ability of organizing conceptually promiscuous data into a uniform structure is critical to knowledge navigating systems. In this paper, we describe a plain indexing method for accumulating diverse information originating from heterogeneous natural information sources. We present a couple of heuristic techniques: (a) *orthogonal decomposition* which allows for reorganizing a given information base into a Cartesian product of primitive concepts, and (b) *analogical refinement* which enables further elaboration based on the measurement of similarity. The proposal is implemented and evaluated against sample information bases containing about a few thousands conceptual units. We discuss the strength and weakness of the method based on the analysis of experimental results.

Introduction

Most of information people produce and manipulate (such as natural language documents, or images) is conceptually promiscuous in the sense that its semantics is not rigorously defined or completely agreed upon. As long term research goals, we aim at (a) designing an inventory of *knowledge media* (Stefik, 1986) with varying degrees of elaboration for accumulating diverse information originating from natural information sources, and (b) constructing a system capable of helping people create, share and explore large knowledge spaces. Towards this end, we are working on the Knowledgeable Community project (Nishida and Takeda, 1993; Nishida *et al.*, 1994; Takeda *et al.*, 1995; Nishida *et al.*, 1995) including several small projects such as information gathering and classification (Iwazume *et al.*, 1995), content-based information summarization (Matsuo *et al.*, 1995), and building a knowledge media information base system called CM-2.

In this paper, we describe a plain indexing method for accumulating diverse information originating from heterogeneous natural information sources. We present a couple of heuristic techniques: (a) *orthogonal decomposition* which allows for reorganizing a given information base into a Cartesian product of primitive concepts, and (b) *analogical refinement* which enables further elaboration based on the measurement of similarity. The proposal is implemented and evaluated against sample information bases containing about a few thousands conceptual units. We discuss the strength and weakness of the method based on the analysis of experimental results.

In what follows, we will first overview CM-2 and describe the role of a plain indexing method using associational representation. We then present heuristic techniques for reorganizing CM-2 information bases. Finally, we show experimental results and make discussion.

Associational Representation in CM-2 Information Base

Architecture of the CM-2 Information Base System

CM-2 is a knowledge media information base system under construction at Nara Institute of Science and Technology.[1] The design goal of CM-2 is to provide users with a means of accumulating, sharing, exploring, and refining conceptually promiscuous information gathered from vast information sources.

CM-2 consists of a collection of *information bases*. Each CM-2 information base is possessed by an individual person or a group (Figure 1). It consists of a collection of *workspaces* and *agents*. Each workspace provides a particular view of looking at information stored in the information base through a particular information representation language. Each agent manipulates information representation or/and interact with the user. The user or agents in CM-2 information bases can interact with other, or incorporate information from other kinds of information sources connected to Internet.

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

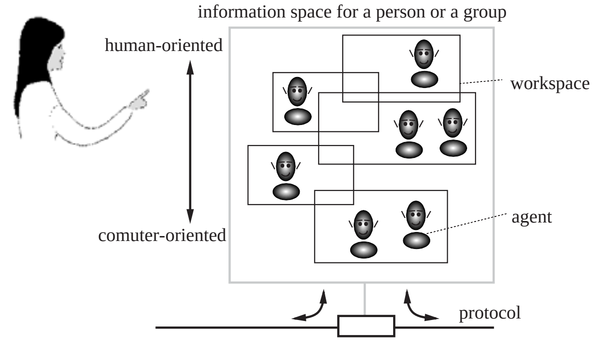

Figure 1: The architecture of the CM-2 information base system

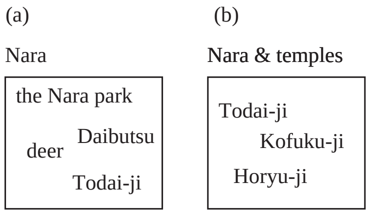

Figure 2: Example of associations in CM-2

Use of Associational Representation in CM-2 Information Base

In this paper, we focus on *associational representation*, which allows the user to explore a way of articulating conceptually promiscuous information by aggregating conceptually relevant information. The basic entities of associational representation are (a) a *unit* which represents either a concept or an external datum, and (b) an *association* which associates a collection of key concepts (hereafter *keys*) with a collection of units (hereafter *values*) which is normally reminded of by the given keys. Figure 2 shows a couple of associations. The left says that given a concept "Nara", one may be reminded of "Todai-ji", "the Nara park", "deer", and "Daibutsu". The right is an example of association with more than one key. It says that "Todai-ji", "Kofuku-ji", and "Horyu-ji" are reminded when "Nara" and "temples" are given as keys.

Conceptual Promiscuity with CM-2 Associational Representation

As suggested by the examples shown in Figure 2, what is reminded of by a given set of keywords may well differ from person to person, and people might give different names to the same concept or the same name to different concepts. We call such a phenomenon *conceptual promiscuity.*

It is our policy to live with conceptual promiscuity, rather than imposing standardization on terminology or conceptualization. Such a "generous mind", as we believe, will be beneficial to us, for it will foster creative thinking. On one hand, we have to admit it because raw data are already there. We would like to incorporate them without spending huge amount of time on reforming them into a well-defined knowledge medium and make use of them before their contents become obsolete. On the other hand, completely well-defined knowledge medium is not possible as philosophers based on the Wittgenstein doctrine have pointed out, for our language can at best be defined connotationally as language games.

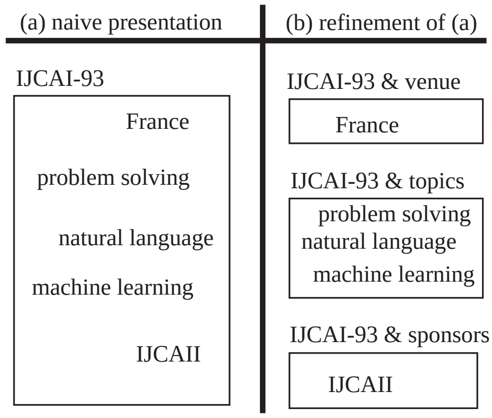

Figure 3: Conceptual promiscuity

In fact, we are experimenting with gathering information from varieties of information sources, such as research notes, WWW pages, and personal memoranda. It is quite easy to generate associational representations from those raw materials and use them as indices.

On the other hand, conceptual promiscuity is not desirable from the information retrieval point of view, for it will cause both noise and incompleteness. Compare two sets of associations in Figure 3. In Figure 3a, the value space associated with the concept "IJCAI-95" is conceptually promiscuous in the sense that entities of various kinds are mixed up there. In contrast, the values of each association in Figure 3b are more coherent. For understanding the domain and utilizing the information, the latter is more useful.[2]

Refining CM-2 Information Base

In order to resolve difficulty with conceptual promiscuity, we explore a way of using heuristics to suggest the user how to refine CM-2 information bases into a coherent structure, as shown in Figure 4.

[2] One may complain about fragmentation of information in Figure 3b and rather prefer the presentation shown in Figure 3a. We cope with such complaints by introducing facilities for aggregating information and present it altogether.

; given an information base B and a threshold $\theta > 0$
repeat
⟨use heuristic rules to diagnose B and produce a set of suggestions and associated penalties if any undesirable portion is found⟩;
if
⟨the largest penalty is greater than θ ⟩;
then
⟨fix B as suggested by the diagnosis with the largest penalty ⟩;
else
exit from the loop
for ever

Figure 4: A general procedure of refinement

In this paper, we assume that people implicitly encode a certain kind of conceptual structure into associational representations even though they are not given an explicit description tool for conceptualization, and we are able to detect such regularity by some means such as statistical analysis.

We present a couple of heuristic techniques which will detect inappropriate associations from CM-2 information base and suggest a possible way of remedying them. *Orthogonal decomposition* attempts to decompose a given information base into coherent groups of associations, by analyzing how the user intersects associations. *Analogical refinement* is a less efficient but more powerful technique for further elaborating information base based on the measurement of similarity. The details will be given in the next two sections.

Orthogonal Decomposition of CM-2 Information Base

Let us first introduce some notations and relations.

Given an association with keys $\{k_1, \ldots, k_m\}$ and values $\{v_1, \ldots, v_n\}$, we denote

$$V[\{k_1, \ldots, k_m\}] = \{v_1, \ldots, v_n\}.$$

Given a set of keys $\{k_1, \ldots, k_m\}$, we define the *extended value* $V^*[\{k_1, \ldots, k_m\}]$ as follows:

$$V^*[\{k_1, \ldots, k_m\}] = \bigcup_{k_{m+1},\ldots,k_n} V[\{k_1, \ldots, k_m\} \cup \{k_{m+1}, \ldots, k_n\}],$$

where $m < n$, $k_i \neq k_j$ whenever $i \neq j$. For example,

$$\begin{aligned} & V^*[\{\text{“IJCAI-93”}\}] \\ &= V[\{\text{“IJCAI-93”}\} \cup \{\text{“venue”}\}] \\ &\quad \cup V[\{\text{“IJCAI-93”}\} \cup \{\text{“topics”}\}] \\ &\quad \cup V[\{\text{“IJCAI-93”}\} \cup \{\text{“sponsors”}\}] \\ &= \{ \text{“France”}, \text{“problem solving”}, \\ &\quad\ \text{“knowledge representation”}, \\ &\quad\ \text{“machine learning”}, \text{“IJCAII”} \} \end{aligned}$$

if
⟨ for concepts x and y:
$V^*[\{x\}] \cap V^*[\{y\}] \neq V^*[\{x,y\}]$ ⟩
then
penalty $\leftarrow \frac{|(V^*[\{x\}] \cap V^*[\{y\}]) - V^*[\{x,y\}]|}{|V^*[\{x,y\}]|}$;
suggestion ← “resolve the difference between $V^*[\{x\}] \cap V^*[\{y\}]$ and $V^*[\{x,y\}]$, by adding z to $V[\{x,y\}]$ if $z \notin V^*[\{x,y\}]$ and $z \in (V^*[\{x\}] \cap V^*[\{y\}])$ ”

if
⟨ for two sets of concepts α, β, $\alpha \subset \beta$:
$\exists z[\, z \in V[\alpha] \wedge z \in V[\beta]\,]$ ⟩
then
penalty $\leftarrow \infty$;
suggestion ← “remove z from $V[\alpha]$.”

Figure 5: Diagnosis rules for orthogonal decomposition

in Figure 3b. Extended values provide a means for looking at the entire picture of entities comprising a given set of keys.

Given a unit x, we define the *extended keys* $K^*[x]$ as

$$K^*[x] = \{\{k_1, \ldots, k_m\} \mid x \in V^*[\{k_1, \ldots, k_m\}]\}.$$

Two set of keys $\{x_1, \ldots, x_m\}$ and $\{y_1, \ldots, y_n\}$ are *orthogonal* to each other (with respect to the associations of information base at that situation), if their associated values, *i.e.*, $V^*[\{x_1, \ldots, x_m\}]$ and $V^*[\{y_1, \ldots, y_n\}]$, have a non-empty intersection. For instance, the key {“IJCAI-93”} is orthogonal to the keys {“venue”}, {“topics”}, and {“sponsors”}, in Figure 3b, for

$$\begin{aligned} & V^*[\{\text{“IJCAI-93”}\}] \cap V^*[\{\text{“venue”}\}] \\ &= \{\text{“France”}, \text{“problem solving”}, \ldots, \} \\ &\quad \cap \{\text{“France”}\} \\ &= \{\text{“France”}\} \neq \{\}. \end{aligned}$$

Orthogonal decomposition is a technique of refining CM-2 information base using diagnosis rules shown in Figure 5.

Example of the application of orthogonal decomposition is illustrated in Figure 6.

Limitation of Orthogonal Decomposition

Orthogonal decomposition is easy and useful, but limited in several ways. First, it cannot make preference. For example, the set of associations in Figure 3b can not be derived from the association shown in Figure 3a even if those shown in Figure 7 are added, for there is no way of deriving the fact that concepts “venue”, “sponsors”, and “topics” are used for referring to more specialized concepts “countries”, “organizations”, and “AI”, respectively.

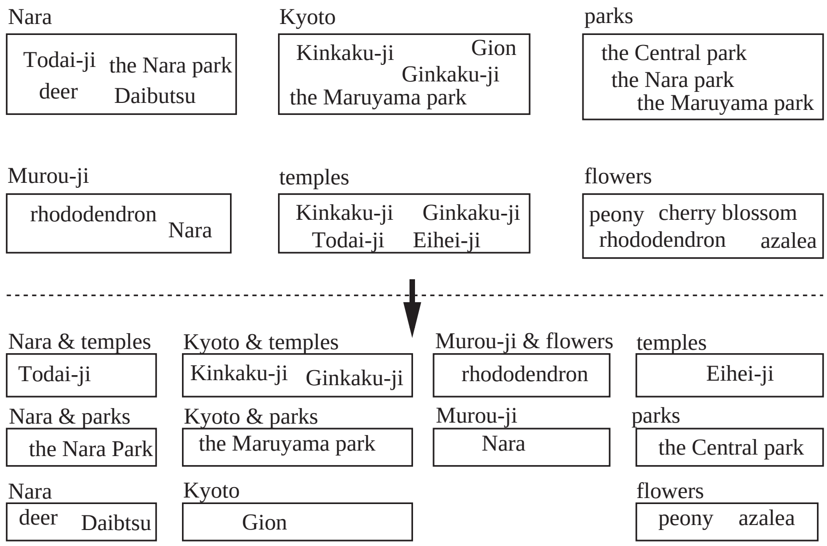

Figure 6: Orthogonal decomposition of CM-2 information base

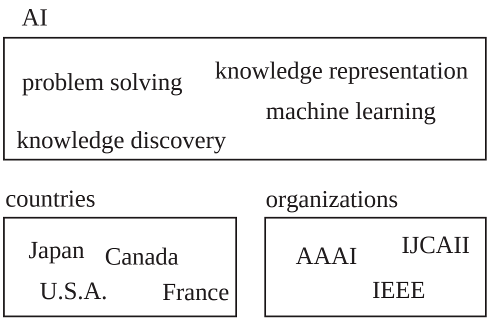

Figure 7: Some associations

Furthermore, orthogonal decomposition may sometimes lead to useless association. For example, although

$$\text{“\$1.00”} \in \mathrm{V}[\{\text{“the fare”}, \text{“the toll bridge A”}]$$

and

$$\text{“\$1.00”} \in \mathrm{V}[\{\text{“the price”}, \text{“the burger”}\}],$$

it is not useful to keep association

$$\begin{aligned}\text{“\$1.00”} \in \mathrm{V}[\{\ & \text{“the fare”}, \text{“the toll bridge A”},\\ & \text{“the price”}, \text{“the burger”}\ \}].\end{aligned}$$

In the next section, we propose a solution for the former problem. The latter will be discussed later in this paper.

Analogical Refinement of CM-2 Information Base

Analogical refinement is based on the measurement of similarity. Given a couple of non-orthogonal keys x and y, we define the similarity $\mathrm{Sim}[x, y]$ as shown in Figure 8. Based on that definition, we define the key similarity $\mathrm{Sim}^*[\alpha, \beta]$ between keys α and β as the sum of maximal pairwise similarities of units in α and β. Namely,

$$\mathrm{Sim}^*[\alpha,\beta] = \max\left[\sum_{x\in\alpha}\max_{y\in\beta}[\mathrm{Sim}[x,y]], \sum_{y\in\beta}\max_{x\in\alpha}[\mathrm{Sim}[x,y]]\right]$$

For concepts x, y, and a threshold $\theta > 0$, we denote $x \sim y$ if $\mathrm{Sim}[x, y] \geq \theta$. Similarly, for keys α, β, and a threshold θ, $\alpha \sim \beta$ if $\mathrm{Sim}^*[\alpha, \beta] \geq \theta$.

The *analogical refinement* heuristic suggests to refine a CM-2 information base according to the following diagnosis rule:

if
  $x \in \mathrm{V}^*[\alpha]$,
  $y \in \mathrm{V}^*[\beta \cup \{a\}]$, and
  $x \notin \mathrm{V}^*[\alpha \cup \{a\}]$
then
  penalty $\leftarrow \mathrm{Sim}[x, y] + \mathrm{Sim}^*[\alpha, \beta]$
  suggestion $\leftarrow$ add x to $\mathrm{V}[\alpha \cup \{a\}]$.

In other words, the above rule will cause the concept x to be added to $\mathrm{V}[\alpha \cup \{a\}]$ if $x \in \mathrm{V}^*[\alpha]$, $\alpha \sim \beta$, and $x \sim y$ for some $y \in \mathrm{V}^*[\beta \cup \{a\}]$.

For example, suppose a set of associations are given as shown in Figure 9, resulting from putting together associations shown in Figure 3b and Figure 7. Consider now the association shown in Figure 10 is added to those shown in Figure 9.

Similarities are computed as partly shown in Table 1 and a new associations will be created as shown in Figure 11. For example, as

Given a couple of non-orthogonal keys x and y, we define the similarity $\mathrm{Sim}[x,y]$ between x and y from three perspectives and let it:

$$\mathrm{Sim}[x,y] = \frac{\mathrm{Sim}^{(a)}[x,y] + \mathrm{Sim}^{(b)}[x,y] + \mathrm{Sim}^{(c)}[x,y]}{3} \in [0,1].$$

$\mathrm{Sim}^{(a)}[x,y]$ measures the similarity between x and y by comparing concepts in $\mathrm{V}^*[\{x\}]$ and those in $\mathrm{V}^*[\{y\}]$. The definition is as follows:

$$\begin{aligned}\mathrm{Sim}^{(a)}[x,y] = \ & \tfrac{1}{|\mathrm{V}^*[\{x\}] \cup \mathrm{V}^*[\{y\}]|} \\ & \cdot(\ |\{z \mid z \in \mathrm{V}^*[\{x\}] \wedge z \in \mathrm{V}^*[\{y\}]\}| \\ & \quad + |\{z \mid z \in \mathrm{V}^*[\{x\}] - \mathrm{V}^*[\{y\}] \wedge \exists u[\, u \in \mathrm{V}^*[\{y\}] \wedge (\mathrm{K}^*[z] \cap \mathrm{K}^*[u] \neq \{\})]\}| \\ & \quad + |\{z \mid z \in \mathrm{V}^*[\{y\}] - \mathrm{V}^*[\{x\}] \wedge \exists u[\, u \in \mathrm{V}^*[\{x\}] \wedge (\mathrm{K}^*[z] \cap \mathrm{K}^*[u] \neq \{\})]\}|).\end{aligned}$$

$\mathrm{Sim}^{(b)}[x,y]$ measures the rate of common keys of associations containing x and y as values. Namely,

$$\mathrm{Sim}^{(b)}[x,y] = \frac{|\{z \mid z \in \mathrm{K}^*[x] \wedge z \in \mathrm{K}^*[y]\}|}{|\mathrm{K}^*[x] \cup \mathrm{K}^*[y]|}.$$

$\mathrm{Sim}^{(c)}[x,y]$ measures the rate of keys orthogonal both to x and to y. Thus,

$$\mathrm{Sim}^{(c)}[x,y] = \frac{|\{z \mid \langle z \text{ is orthogonal to } x\rangle \wedge \langle z \text{ is orthogonal to } y\rangle\}|}{|\{z \mid \langle z \text{ is orthogonal to } x\rangle\} \cup \{z \mid \langle z \text{ is orthogonal to } y\rangle\}|}.$$

Figure 8: Defining similarity between concepts

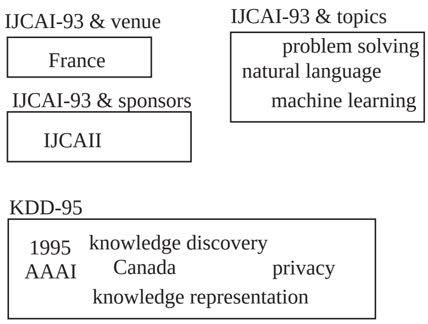

Figure 9: The set of associations resulting from putting together associations shown in Figure 3b and Figure 7

$$\begin{aligned}\mathrm{K}^*[\text{“Canada”}] &= \{\{\text{“countries”}\}, \{\text{“KDD-95”}\}\} \\ \mathrm{K}^*[\text{“Japan”}] &= \{\{\text{“countries”}\}\}\end{aligned}$$

so,

$$\begin{aligned}&\mathrm{Sim}^{(b)}[\text{“Canada”}, \text{“Japan”}] \\ &= \frac{|\{\{\text{“countries”}\}\}|}{|\{\{\text{“countries”}\}, \{\text{“KDD-95”}\}\}|}\end{aligned}$$

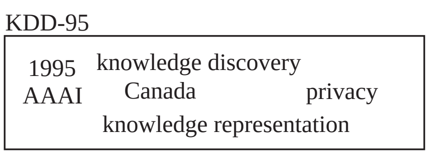

Figure 10: An association added to the associations shown in Figure 9

Table 1: Part of similarities computed for associations shown in Figure 10, Figure 3b, and Figure 7

x	y	$\mathrm{Sim}[x,y]$
“Canada”	“Japan”	$\frac{1}{3} \cdot \frac{1}{2}$
“France”	“U.S.A.”	$\frac{1}{3} \cdot \frac{1}{2}$
“knowledge representation	“problem solving”	$\frac{1}{3} \cdot \frac{1}{3}$
“knowledge discovery”	“machine learning”	$\frac{1}{3} \cdot \frac{1}{3}$
“AAAI”	“IJCAII”	$\frac{1}{3} \cdot \frac{1}{3}$
“KDD-95”	“IJCAI-93”	$\frac{1}{3} \cdot \frac{8}{11}$

$$= \frac{1}{2}.$$

On the other hand, as

$$\mathrm{V}^*[\text{“Canada”}] = \mathrm{V}^*[\text{“Japan”}] = \{\}$$

and no keys have intersection with keys $\{$“Canada”$\}$ or $\{$“Japan”$\}$, so

$$\begin{aligned}&\mathrm{Sim}^{(b)}[\text{“Canada”}, \text{“Japan”}] \\ &= \mathrm{Sim}^{(c)}[\text{“Canada”}, \text{“Japan”}] \\ &= 0.\end{aligned}$$

Hence,

$$\begin{aligned}&\mathrm{Sim}[\text{“Canada”}, \text{“Japan”}] \\ &= \frac{1}{3} \cdot \mathrm{Sim}^{(a)}[\text{“Canada”}, \text{“Japan”}] \\ &= \frac{1}{3} \cdot \frac{1}{2}.\end{aligned}$$

Using the similarities,

$$\begin{aligned}\text{“KDD-95”} &\sim \text{“IJCAI-93”} \\ \text{“knowledge representation”} &\sim \text{“problem solving”}\end{aligned}$$

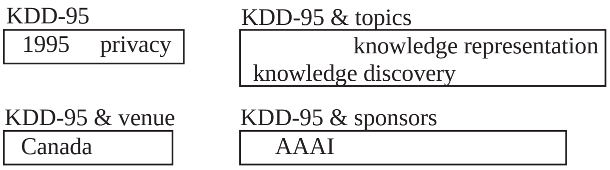

Figure 11: New contexts resulting from adding a context shown in Figure 10 to those shown in Figure 3b and Figure 7

we will obtain

$$\text{“knowledge representation”} \in \mathrm{V}[\{\text{“KDD-95”}, \text{“topics”}\}]$$

The results obtained are conceptually more well-structured than those obtained solely by orthogonal decomposition.

Experiments

We have implemented a kernel of CM-2 on top of Common Lisp and tcl/tk. We are evaluating CM-2 against accumulating various kinds of information such as research memoranda, technical surveys, regional guide, personal diary, and so on. Besides testing against these small examples and the examples described so far, we have made a couple of experiments with a nontrivial scale.

Experiment 1 We have manually constructed a CM-2 information base for regional guide of Nara, Japan. It contains about 1,850 concepts and 870 associations. As a result of orthogonal decomposition, CM-2 has produced 212 revisions, about 80 of which have been found useful. Others are uninteresting. On the other hand, the analogical refinement heuristic has generated 65 suggestions, 20 of which are found useful.

There are several interesting suggestions. For example, from “cherry blossom” $\in$ V[{“Ikoma park”}] and, “cherry blossom” $\in$ V[{“flowers”}], we obtained

$$\text{“cherry blossom”} \in \mathrm{V}[\{\text{“Ikoma park”}, \text{“flowers”}\}],$$

from which we in turn obtained

$$\text{“iris”} \in \mathrm{V}[\{\text{“Ayameike park”}, \text{“flowers”}\}]$$

based on

$$\begin{array}{rcl} \text{“iris”} & \in & \mathrm{V}[\{\text{“Ayameike park”}\}], \\ \text{“Ikoma park”} & \sim & \text{“Ayameike park”, and} \\ \text{“cherry blossom”} & \sim & \text{“iris”}, \end{array}$$

as shown in Figure 12.

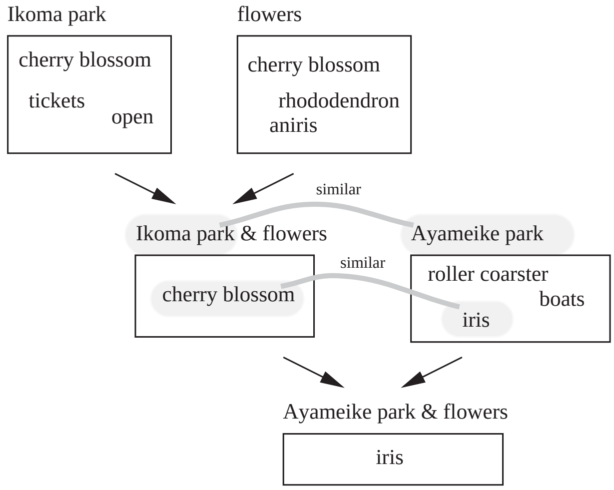

Figure 12: An interesting suggestion obtained by orthogonal decomposition and analogical refinement

Experiment 2 We have gathered WWW pages in the Kansai (west) area of Japan and have generated associational representations using a simple keyword extraction and text analysis program. As a result, we have obtained about 700 concepts and 230 associations. CM-2 has proposed 1130 and 230 revisions using the orthogonal decomposition and analogical refinement heuristics. About 230 and 12 proposals by orthogonal decomposition and analogical refinement heuristics, respectively, are found useful.

Discussion

This work is complementary to recent work on information gathering from heterogeneous sources on Internet (Levy *et al.*, 1994; Armstrong *et al.*, 1995; Balabanovi'c and Shoham, 1995; Li, 1995). 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.

The basic recognition behind this research is a trade-off between the benefit from conceptually well-structured information space and the cost for organizing information space. The more organized is the information space, the more utilities can a navigator provide, while the higher cost we have to pay. When we look at vast information from various information sources accessible through a global computer network, it seems almost impossible to design a well-defined conceptual structure on which information is accumulated.

Our approach is to provide a framework of information representation with a low structural facilities and to facilitate raw information from vast information sources to be incorporated without much labor and gradually refined and elaborated as more insights are obtained.

How successful is our approach? Our early experiments have ended up in very promising results as far as a half of our goal is concerned. Members of our group have been able to use associational representation to accumulate varieties of information taken from vast information sources and access relevant information.

However, it turned out that there is much space to improve as for the other half of our goal: to use heuristics to semi-automatically introduce conceptual structure into the information base.[3] Unfortunately, the rate of useful suggestions from the heuristics, being less than 30%, seems to be too low, though we have not evaluated how well the given set of sample associations is conceptually structured from the beginning.

To improve the quality of heuristics, we are currently looking at introduction of other kinds of heuristics and a domain knowledge.

One possibility is introducing a notion of *significance* of association. Given a set of keys α, we may ask the user to specify either α is *significant* or not (*i.e.*, *insignificant*), depending on whether thinking about α makes sense to the user or not, respectively. If the user has explicitly given the contents of $\mathrm{V}^*[\alpha]$, α is taken to be significant. The user may well reserve the remark about the significance of the keys α. In such a case, α is called *significance-unknown*. Using this convention, we can avoid thinking about such associations as

$$\begin{aligned}\text{“\$1.00”} \in \mathrm{V}[\{\ & \text{“the fare”}, \text{“the toll bridge A”}, \\ & \text{“the price”}, \text{“the burger”}\}].\end{aligned}$$

Or, we might use the following kind of heuristic and store “hindsight” as a more structured conceptual structure.

$$\begin{array}{l}\textbf{if} \\ \quad \langle\ \mathrm{V}^*[\{x,y\}] \text{ does not make sense, or} \\ \quad \text{the user assert that} \\ \qquad \mathrm{V}^*[\{x,y\}] \neq (\mathrm{V}^*[\{x\}] \cap \mathrm{V}^*[\{y\}])\ \rangle \\ \textbf{then} \\ \quad \forall z[\ \ z \in (\mathrm{V}^*[\{x\}] \cap \mathrm{V}^*[\{y\}]) \\ \qquad \rightarrow \neg\langle z \ isa\ x\rangle \wedge \neg\langle z\ isa\ y\rangle\ ].\end{array}$$

Further research is open for future.

Conclusion

In this paper, we have proposed an approach to accumulating diverse information using a plain indexing method. We have given a couple of heuristic techniques, orthogonal decomposition and analogical refinement, for refining the conceptual structure of information base. We have discussed the strength and weakness of the method based on the analysis of experimental results.

[3] Huffman and Steier (Huffman and Steier, 1995) propose a similar method using heuristic join to combine data from multiple structured sources.

References

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

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

Huffman, Scott B. and Steier, David 1995. Heuristic joins to integrate structured heterogeneous data. In *Working Notes of the AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments.* 74–77.

Iwazume, Michiaki; Takeda, Hideaki; and Nishida, Toyoaki 1995. Ontology-based approach to information gathering and text categorization. (to be presented at ISDL'95).

Levy, Alon Y.; Sagiv, Yehoshua; and Srivasava, Divesh 1994. Towards effcient information gathering agents. In *Working Notes of the AAAI Spring Symposium on Software Agents.* 64–70.

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

Matsuo, Toshiyuki; Takeda, Hideaki; and Nishida, Toyoaki 1995. An application of knowledge medium to semi-automated summarization of technical papers in metallurgy. In *JSAI Annual Convention Records.* 15-07 (in Japanese).

Nishida, Toyoaki and Takeda, Hideaki 1993. Towards the knowledgeable community. In *Proceedings International Conference on Building and Sharing of Very-Large Scale Knowledge Bases '93 (KB&KS '93).* Ohmsha, Ltd. 157–166.

Nishida, Toyoaki; Takeda, Hideaki; and Iino, Kenji 1994. Mediation in the knowledgeable community. In *Working Notes of the AAAI Spring Symposium on Software Agents.*

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

Stefik, Mark 1986. The next knowledge medium. *AI Magazine* 7(1):34–46.

Takeda, Hideaki; Iino, Kenji; and Nishida, Toyoaki 1995. Ontology-supported agent communication. In *Working Notes of the AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments.* AAAI. 160–165.