

An Associative Search Method Based on Symbolic Filtering and Semantic Ordering for Database Systems

*Naofumi Yoshida**, *Yasushi Kiyoki*** and *Takashi Kitagawa**

** Institute of Information Sciences and Electronics
University of Tsukuba
Tsukuba, Ibaraki 305, Japan
phone: +81-298-53-5163, fax: +81-298-53-5206
e-mail: naofumi@softlab.is.tsukuba.ac.jp,
takashi@is.tsukuba.ac.jp*

*** Faculty of Environmental Information
Keio University
Fujisawa, Kanagawa 252, Japan
phone: +81-466-47-5111, fax: +81-466-47-5041
e-mail: kiyoki@sfc.keio.ac.jp*

Abstract

Associative search is a key function for extracting information from databases. In this paper, we present a new associative search method with symbolic and semantic operations. This method integrates two kinds of associative search functions. The symbolic associative search is a simple pattern-matching-based function. This function is used as the information filter which repeatedly executes pattern-matching-based comparisons between data items. The semantic associative search function extracts semantically related information by mathematical semantic operations based on the mathematical model of meaning which we have proposed. This function provides a context recognition mechanism for extracting semantically related information from databases. This mechanism makes it possible to put the semantically related data items in order, according to the correlation to the searcher's impression. The integrated associative search method realizes the advanced information extraction by combining the symbolic and semantic associative search functions.

Keywords

semantic associative search, database system, image database, mathematical model of meaning

1 INTRODUCTION

Databases have been widely spread in world-wide computer networks. The main operation for information extraction from databases is associative search. A symbolic associative search function is widely used as a method of associative search. It is difficult for this function to extract semantically related information which has the same or similar meaning with different representations (Kolodner 1984, Krikelis et al. 1994, Potter 1992).

A number of associative search methods have been proposed for realizing efficient information extraction in database and knowledge base systems (David et al. 1982, Krikelis et al. 1994, Potter 1992). In the previously proposed methods, relationships between data items are represented explicitly by using information of connections. The relationships are extracted by using simple pattern matching operations and pursuing the information of the connection (Kolodner 1984). In those methods, from the view point of semantic representation of data items, the meanings of the data items are definitely fixed, and those data items are used as information with the fixed meanings. We consider that the relationships between data items vary in the response of situations or contexts. As a model for measuring the semantic relationship dynamically between data items with recognizing the context, we have designed a mathematical model of meaning (Kitagawa et al. 1993, Kiyoki et al. 1994, Kiyoki et al. 1995).

The symbolic associative search function is effective when the symbolic patterns of retrieval target data items are unambiguously defined. This function is widely used in information extraction in database systems. The semantic associative search function by the mathematical model of meaning dynamically computes semantic correlations between a given context and retrieval candidate data items by semantic operations. This function is used for extracting semantically related information to a context given by a searcher. The main feature of the mathematical model of meaning is that the semantic associative search is performed in the orthogonal semantic space. This space is created for dynamically computing semantic correlations between the given context and retrieval candidate data items.

In this paper, we present a new associative search method with the integrated functions of the symbolic and semantic associative search. In this method, appropriate data items are filtered by the pattern-matching-based symbolic associative search function and those data items are semantically put in order by the semantic associative search function. This method realizes highly functional information extraction from databases. We have designed this system as a heterogeneous information processing system in a

multidatabase environment (Bright et al. 1992, Litwin et al. 1990, Sheth et al. 1990). We have designed an integration function for combining the symbolic associative search function and the semantic associative search function in the metalevel architecture. By introducing this function, the functional integration between symbolic and semantic associative search functions is realized without modifying the implementations of those functions.

Several information retrieval methods, which use the orthogonal space created by the mathematical procedures like SVD (Singular Value Decomposition), have been proposed (e.g. the Latent Semantic Indexing method (Deerwester et al. 1990)). Our semantic associative search method is essentially different from those methods using the SVD. The essential difference is that our method provides the important function for semantic projections which realizes the dynamic recognition of contexts. That is, in our method, the context-dependent interpretation is dynamically performed for computing the correlations between a given context and retrieval candidate data items by selecting a subspace from the entire orthogonal semantic space. In our method, the number of phases of the contexts is almost infinite (currently 2^{800} , approximately). Other methods do not provide the context dependent interpretation for computing equivalence and similarity in the orthogonal space, that is, the phase of meaning is fixed and static in those methods.

To compute the semantic relationships between data items, several fuzzy relational database systems have been proposed (Raju et al. 1988, Rundensteiner 1989). In comparison to the fuzzy relational database systems, the essential difference of our system is that our system eliminates ambiguity in semantic operations by introducing the concept on context recognition.

The mathematical model of meaning is a new model for realizing the semantic associative search and extracting semantically related information by giving context words. This model can be applied to extract media data items by giving the context words which represent the impression and contents of the media data items (Kiyoki et al. 1994).

The mathematical model of meaning consists of:

- 1) A set of m words is given, and each word is characterized by n features. That is, m by n matrix is given as the data matrix.
- 2) The correlation matrix with respect to the n features is constructed. Then, the eigenvalue decomposition of the correlation matrix is computed and the eigenvectors are normalized. The orthogonal semantic space is created as the span of the eigenvectors which correspond to nonzero eigenvalues.
- 3) Media data items and context words are characterized by using the specific features (words) and representing them as vectors (The sequence of context words are used to represent a context).
- 4) The media data items and context words are mapped into the orthogonal semantic space by computing the Fourier expansion for the vectors.
- 5) A set of all the projections from the orthogonal semantic space to the invariant sub-

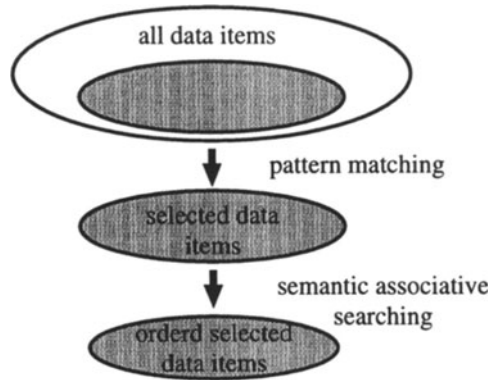


Figure 1 An overview of the associative search process

spaces (eigen spaces) is defined. Each subspace represents a phase of meaning, and it corresponds to a context or situation.

6) A subspace of the orthogonal semantic space is selected according to the user's impression or the content of media data items, which is given as a context represented by a sequence of words.

7) The closest media data item to the context representing in the user's impression and the contents of media data items is extracted in the selected subspace.

2 THE ASSOCIATIVE SEARCH PROCESSES

2.1 Associative search procedure

The associative procedure consists of two steps as shown in Figure 1.

Step 1 : Filtering the data items by the symbolic associative search.

Step 2 : Ordering the selected data items by the semantic associative search.

To realize these steps, our method provides three basic functions.

In this method, it is assumed that each data item is identified by an identifier commonly shared between the symbolic and the semantic associative search functions (Figure 2). In the symbolic associative search function, the metadata items of the media data (e.g. media name, authors, created date) are represented in the form of the relational database. In these semantic associative search function, vectors corresponding to media data items(media data vectors) are stored on the semantic space.

2.2 Function-1: symbolic associative search

It is the best way for performing symbolic associative search to use a relational database system. By using the selection operation of the relational database system, we can select the data items with the same pattern as a keyword and obtain a set of identifiers of the selected data items.

2.3 Function-2: semantic associative search

In the mathematical model of meaning, semantic associative search is to extract the data item with the highest correlation to a given context from the specific data item set. The procedures of these steps are as follows:

Step-1 : Context recognition

Given a sequence of context words for determining the context, the recognition of the context is performed by the method described in the Section 3.

Step-2 : Ordering of the data items:

The data item with the highest correlation to the given context is selected from the retrieval candidate data items. That is, according to the given context, the data item with the highest correlation to the context words is selected from the retrieval candidate data item set. This selection is repeatedly performed, and the data items are put in order, according to their correlations to the given context.

2.4 Function-3: integration of associative search functions

This integration function is used for integrating symbolic and semantic associative search functions indirectly. This function enables the symbolic and semantic associative search functions to be combined through this function as the meta-level function.

The procedure is as follows:

Step-1 : Selection by the symbolic associative search:

The data items in the source relation are filtered by the symbolic associative search function. As the result, the integration function obtains the identifiers of the selected data items. The integration function gives the identifiers of selected data items to the semantic associative search function.

Step-2 : Ordering by the semantic associative search:

The semantic associative search function receives the set of the identifiers. The identifiers are ordered by the semantic associative search function and

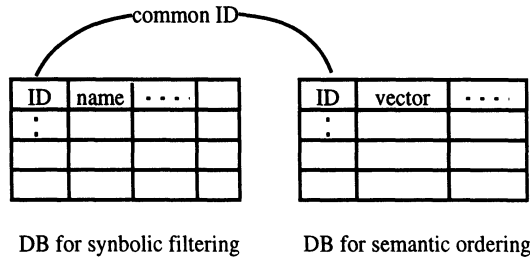


Figure 2 Commonly shared identifiers

the ordered list is returned to the integration function. The integration function orders and outputs the tuples of the relation according to the ordered identifiers obtained by the semantic associative search function.

3 FORMALIZATION OF SEMANTIC ASSOCIATIVE SEARCH

In this section, we review the mathematical model of meaning (Kitagawa et al. 1993, Kiyoki et al. 1994, Kiyoki et al. 1995), which is the basic model for the semantic associative search function.

3.1 Creation of a semantic space

The semantic associative search is realized by the mathematical model of meaning (Kitagawa et al. 1993, Kiyoki et al. 1994, Kiyoki et al. 1995) which we have proposed. For the data items for space creation, a data matrix M is created. When m data items for space creation are given, each data item is characterized by n features (f_1, f_2, \dots, f_n) . For given $\mathbf{d}_i (i = 1, \dots, m)$, the data matrix M is defined as the $m \times n$ matrix whose i -th row is \mathbf{d}_i . Then, each column of the matrix is normalized by the 2-norm in order to create the matrix M .

Figure 3 shows the matrix M . That is $M = (\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3, \dots, \mathbf{d}_n)^T$.

1. The correlation matrix $M^T M$ of M is computed, where M^T represents the transpose of M .

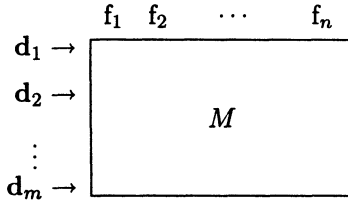


Figure 3 Representation of semantic items by matrix M

2. The eigenvalue decomposition of $M^T M$ is computed.

$$M^T M = Q \begin{pmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_\nu & \\ & & & 0 \dots 0 \end{pmatrix} Q^T,$$

$$0 \leq \nu \leq n.$$

The orthogonal matrix Q is defined by

$$Q = (\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n)^T,$$

where \mathbf{q}_i 's are the normalized eigenvectors of $M^T M$. We call the eigenvectors "semantic elements" hereafter. Here, all the eigenvalues are real and all the eigenvectors are mutually orthogonal because the matrix $M^T M$ is symmetric.

3. Defining the semantic space \mathcal{MDS} .

$$\mathcal{MDS} := \text{span}(\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_\nu),$$

which is a linear space generated by linear combinations of $\{\mathbf{q}_1, \dots, \mathbf{q}_\nu\}$. We note that $\{\mathbf{q}_1, \dots, \mathbf{q}_\nu\}$ is an orthonormal basis of \mathcal{MDS} .

3.2 The set of the semantic projections Π_ν

The projection P_{λ_i} is defined as follows:

$P_{\lambda_i} \stackrel{d}{\iff}$ Projection to the eigenspace corresponding to the eigenvalue λ_i ,
i.e. $P_{\lambda_i} : \mathcal{MDS} \rightarrow \text{span}(\mathbf{q}_i)$.

The set of the semantic projections Π_ν is defined as follows:

$$\Pi_\nu :=$$

$$\begin{aligned} & \{ 0, P_{\lambda_1}, P_{\lambda_2}, \dots, P_{\lambda_\nu}, \\ & P_{\lambda_1} + P_{\lambda_2}, P_{\lambda_1} + P_{\lambda_3}, \dots, P_{\lambda_{\nu-1}} + P_{\lambda_\nu}, \\ & \quad \vdots \\ & P_{\lambda_1} + P_{\lambda_2} + \dots + P_{\lambda_\nu} \}. \end{aligned}$$

The number of the elements of Π_ν is 2^ν , and accordingly it implies that 2^ν different phases of meaning can be expressed by this formulation.

3.3 Semantic operator

The correlations between each context word and each semantic element are computed by this process. The context word is used to represent the user's impression and the contents for media data items to be extracted. A sequence

$$s_\ell = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_\ell)$$

of ℓ context words and a positive real number $0 < \varepsilon_s < 1$ are given, the semantic operator S_p constitutes a semantic projection $P_{\varepsilon_s}(s_\ell)$, according to the context. That is,

$$S_p : T_\ell \mapsto \Pi_\nu$$

where T_ℓ is the set of sequences of ℓ words and $T_\ell \ni s_\ell$, $\Pi_\nu \ni P_{\varepsilon_s}(s_\ell)$. Note that the set $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_\ell\}$ must be a subset of the words defined in the matrix M .

The constitution of the operator S_p consists of the following processes:

1. Fourier expansion of \mathbf{u}_i ($i = 1, 2, \dots, \ell$).
The inner product of \mathbf{u}_i and \mathbf{q}_j u_{ij} is computed, i.e.

$$u_{ij} := (\mathbf{u}_i, \mathbf{q}_j), \text{ for } j = 1, 2, \dots, \nu.$$

We define $\hat{\mathbf{u}}_i \in \mathcal{I}$ as

$$\hat{\mathbf{u}}_i := (u_{i1}, u_{i2}, \dots, u_{i\nu}).$$

This is the mapping of the context word \mathbf{u}_i to the semantic space \mathcal{MDS} .

2. Computing the semantic center $\mathbf{G}^+(s_\ell)$ of the sequence s_ℓ .

$$\mathbf{G}^+(s_\ell) := \frac{\left(\sum_{i=1}^{\ell} u_{i1}, \dots, \sum_{i=1}^{\ell} u_{i\nu} \right)}{\left\| \left(\sum_{i=1}^{\ell} u_{i1}, \dots, \sum_{i=1}^{\ell} u_{i\nu} \right) \right\|_{\infty}},$$

where $\|\cdot\|_{\infty}$ denotes infinity norm.

3. Determining the semantic projection $P_{\varepsilon_s}(s_\ell)$.

If the sum for a semantic element is greater than a given threshold ε_s , we employ the semantic element to form the projected semantic subspace. We define the semantic projection by the sum of such projections.

$$P_{\varepsilon_s}(s_\ell) := \sum_{i \in \Lambda_{\varepsilon_s}} P_{\lambda_i} \in \Pi_{\nu},$$

where $\Lambda_{\varepsilon_s} := \{ i \mid |(\mathbf{G}^+(s_\ell))_i| > \varepsilon_s \}$.

3.4 The Creation method of metadata for media data items

The media data item (e.g. image data item) P consists of t objects (or impression words) $\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_t$, where each object is defined as an n dimensional vector:

$$\mathbf{o}_i = (o_{i1}, o_{i2}, \dots, o_{in}),$$

which is characterized by t specific features.

Namely, we define the media data item P as the collection of t objects (or impression words).

$$P = \{\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_t\}.$$

Moreover, we define the operator union \bigoplus of objects $\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_t$, to represent the metadata for the media data items P as a vector as follows:

$$\bigoplus_{i=1}^t \mathbf{o}_i \equiv (\text{sign}(o_{\ell_1 1}) \max_{1 \leq i \leq t} |o_{i1}|, \text{sign}(o_{\ell_2 2}) \max_{1 \leq i \leq t} |o_{i2}|, \dots, \text{sign}(o_{\ell_n n}) \max_{1 \leq i \leq t} |o_{in}|),$$

where $\text{sign}(a)$ represents the sign (plus or minus) of “ a ” and $\ell_k, k = 1, \dots, n$, represents the index which gives the maximum, that is:

$$\max_{1 \leq i \leq t} |o_{ik}| = |o_{\ell_k k}|.$$

3.5 The expression for semantic associative search

We introduce an expression to measure the correlation between the context words (keywords) and the media data items. This expression measures the correlation between a set of context words and each retrieval candidate media data item.

We can regard the set of context words as the words forming the context s_ℓ . We can specify semantic subspaces with weights c_j 's. Since the norm of a media data item, which can be calculated from the metadata of the media data item, reflects the correlation between the media data item and the semantic elements included in each selected subspace, we may use it as the measure for the correlation between the given context and each media data item.

The expression $\bar{\eta}_0(\mathbf{x}; s_\ell)$ for computing the norm of a media data item, in which we eliminates the effect of the negative correlation by omitting the corresponding terms, is defined as follows:

$$\bar{\eta}_0(\mathbf{x}; s_\ell) = \frac{\sqrt{\sum_{j \in \Lambda_{e_s} \cap S} \{c_j(s_\ell)x_j\}^2}}{\|\mathbf{x}\|_2},$$

where the set S is defined by $S = \{i | \text{sign}(c_i(s_\ell)) = \text{sign}(x_i)\}$ and the weight $c_j(s_\ell)$ is given as follows:

$$c_j(s_\ell) := \frac{\sum_{i=1}^{\ell} u_{ij}}{\left\| \left(\sum_{i=1}^{\ell} u_{i1}, \dots, \sum_{i=1}^{\ell} u_{i\nu} \right) \right\|_{\infty}}, \quad j \in \Lambda_{e_s}.$$

3.6 The semantic associative search algorithm

The semantic associative search function realizes context-dependent interpretation. This function performs the selection of the semantic subspace from the semantic space (\mathcal{MDS}). When a sequence s_ℓ of context words for determining a context are given to the system, the selection of the semantic subspace is performed. This selection corresponds to the recognition of the context, which is defined by the given context words. The selected semantic subspace corresponds to a given context. In the selected semantic subspace, the media data item with the highest correlation to the given context is obtained by the expression $\bar{\eta}_0(\mathbf{x}; s_\ell)$ defined in Section 3.5.

This semantic associative search is performed by the following procedure:

- Step-1** When a sequence s_ℓ of the context words for determining a context (representing the user's impression and the contents of media data item) are given, the Fourier expansion is computed for each context word, and the Fourier coefficients of these words with respect to each semantic element are obtained. This corresponds to seeking the correlation between each context word and each semantic element.
- Step-2** The values of the Fourier coefficients for each semantic element are summed up to find the correlation between the given context words and each semantic element.
- Step-3** If the sum obtained in Step-2 in terms of each semantic element is greater than a given threshold ε_s , the semantic element is employed to form the semantic subspace $P_{\varepsilon_s}(s_\ell)MDS$. This corresponds to the recognition of the context.
- Step-4** By using the expression $\bar{\eta}_0(\mathbf{x}; s_\ell)$, the metadata item for the media data item with the highest correlation to the context is selected among the candidate metadata items for the media data set in the selected semantic subspace. This corresponds to finding the media data item with the highest correlation of the given context.

4 DATA STRUCTURES AND PRIMITIVE OPERATORS

In this section, we describe the data structures and the primitive operators for implementing our method.

4.1 Data structures and primitive operators in the symbolic associative search function

In the symbolic associative search function, the data structure and several primitive operators of the relational database system are used. We use the "relation" as the data structure.

We define the set of primitive operators in the symbolic associative search as follows:

- (select [rel] [att] [cond] [val])
- (project [rel] [att-list])
- (join [rel1] [att1] [rel2] [att2] [cond])
- (union [rel1] [rel2])
- (diff [rel1] [rel2])

(The parameters "rel, rel1, rel2" : relations, "att, att1, att2" : at-

tributes, “att-list” : a list of attributes, “cond” : a condition, and “val” : a keyword value.)

4.2 Data structures and primitive operators of the semantic associative search function

In this section, we describe the data structure and the primitive operator on the semantic associative search function.

In the semantic associative search function, a vector corresponds to a media data item. In the primitive operator of semantic associative search, a context, a set of vectors of retrieval candidate media data items, and the number of return values are given as parameters.

We define the primitive operator as follows:

- (semantic-search [context] [target] [maxresult])

(The parameter “context” : a context, “target” : a set of retrieval candidate vectors, and “maxresult” : the number of return values.)

4.3 Data structures and primitive operators in the associative search method

In this section, we describe the data structure and the primitive operator in the integrated system.

As the data structure, we use “relation” in the symbolic associative search function and a set of vectors of retrieval candidate media data in the semantic associative search function. As the assumption, the common identifiers are shared to recognize the same objects between those functions.

This primitive operator is the higher level operator of the semantic associative search function (in the previous section). The parameters of the primitive operator are a relation, context, a set of vectors of retrieval target media data, a semantic space for semantic associative search, the number of results n , and names of the attributes which are added. The return value of the primitive operator is the relation with the added attributes. The attributes of the ranking and the norms are included.

- (search-mediadata-by-context [rel] [ID]
[att1] [att2] [space] [target] [user]
[maxresult] [context])

(The parameters “rel” : an input relation, “ID” : an attribute name of

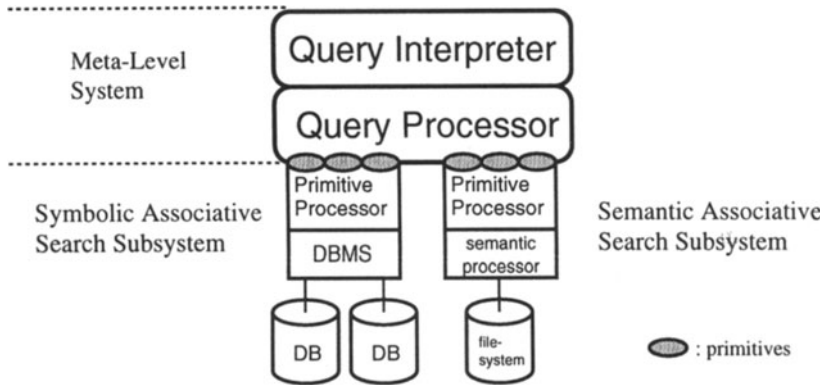


Figure 4 The implementation of the experimental system

the identifier, “att1, att2” : names of the attributes, “space” : a semantic space, “target” : a set of vectors of retrieval target media data, “user” : a user’s dictionary, “maxresult” : the number of results, and “context” : a sequence of context words.)

The procedure of this primitive operator is as follows:

- Step-1** Measures correlations in the semantic space by Function-2 of Section 2.3 with a given context, a set of vectors of retrieval candidate media data, a semantic space, and the number of results n .
- Step-2** Orders the media data items according to the correlation, and obtains the top n data items (The ranking and norms of the n vectors are obtained).
- Step-3** Adds attributes of the ranking and norms in the input relation.
- Step-4** Inserts the attribute values of ranking and norms into the input relation.

5 IMPLEMENTATION OF THE SYSTEM

In this section, we present the implementation of the proposed associative search method.

We have implemented the associative search method by the system consisting of three modules (Figure 4). They are named the symbolic associative search subsystem, the semantic associative search subsystem, and the meta-level system.

5.1 Implementation of the symbolic associative search subsystem

We have implemented the symbolic associative search (pattern matching) method by using the relational database system. This subsystem is implemented with the primitive processor and the database management system (DBMS), as shown in Figure 4.

The primitive processor is the mediator between the meta-level system and the DBMS. By the uniform access through primitives, our system realizes the abstraction of multiple DBMSs. Relations are used as the data structure.

5.2 Implementation of semantic associative search subsystem

We have implemented an experimental system of the semantic associative search function. To create a data matrix M automatically, we have referred to the English dictionary named "General Basic English Dictionary (Ogden 1940)" in which only 871 basic words are used to explain every English vocabulary entry. Those basic words are used as features, that is, they are used as the features corresponding to the columns in the data matrix M . Namely, 871 features are provided to make the semantic space. And, 2115 words are used to represent the words corresponding to the rows in the data matrix M . These 2115 words have been selected as the basic vocabulary entries. These entries are the same as the basic explanatory words used in the English dictionary named "Longman Dictionary of Contemporary English (Longman 1987)." The 2115×871 data matrix is used to create the semantic space.

By using this matrix, the semantic space is created. Context words and retrieval candidate media data items are mapped into this space. Furthermore, each basic word corresponding to vocabulary entries is mapped to the semantic space by the Fourier expansion. The procedure for creation of the semantic space is as follows:

1. Each of 2115 vocabulary entries corresponds to a row of the matrix M . In the setting of a row of the matrix M , each column corresponding to the explanatory words (features) which appear in each vocabulary entry is set to the value "1". If the explanatory word is used as the negative meaning, the column corresponding to the word (feature) is set to the value "-1". The column corresponding to the vocabulary entry itself is set to the value "1". And, the other columns are set to the value "0". This process is performed for every vocabulary entry. And then, each column of the matrix is normalized by the 2-norm to create the matrix M .
2. By using this matrix M , an semantic space is computed as described in Section 3.

To create the data matrix M from the dictionary automatically, we have implemented several filters which remove unnecessary words, such as articles and pronouns, and transform conjugations and inflections of words to the infinitives. The unnecessary words are not used as features in the data matrix M .

Each English word is mapped into this semantic space. We used the simple words which appear in the dictionary itself. We have performed several experiments using this semantic space to clarify the effectiveness of our method.

5.3 Implementation of the meta-level system

We have implemented the meta-level system consisting of the query interpreter and the query processor (Figure 4). The query interpreter receives a searcher's (user's) query and translates it into the sequence of primitives. And then, it sends the sequence to the query processor. The query processor distributes primitives to the subsystems.

The query processor is currently implemented on the UniSQL(UniSQL 1995), which is the extended system on object-orientation in the relational database system.

6 EXPERIMENTS

We performed several experiments to clarify the feasibility of our associative search method. We have made it clear that the method based on symbolic filtering and semantic ordering realizes the advanced associative search for media data.

6.1 Experimental environment

We have implemented the experimental system in C language and ESQL/X language(UniSQL 1995). As the platform, we have used Sun SparcStation 5 and Sun SparcStation EC (SunOS 4.1.4).

6.2 Experiment-1

In this experiment, the 30 image data items are used as media data. The image database is named "famousimages" and listed as shown in Figure 5. The metadata (impression words) of these images are given as shown in Figure 6. The query used in Experiment-1 is shown in Figure 7. The data items with "Hokusai" as the author are selected by the selection operation. And then,

those data items are ordered, according to the given context “power, force”, by the semantic associative search subsystem. This query means “Select the Hokusai’s images, and put them in order, according to the context ‘power’ and ‘force’.” The results are listed as shown in Figure 8.

6.3 Experiment-2

As Experiment-2, the query is shown in Figure 9. This query means “Select images which were painted after 1950, and put them in order, according to the context ‘light’ and ‘bright’.” By the selection operation, the images painted after 1950 are selected, and then, they are ordered, according to the context “light, bright”, by the semantic associative search function. The results are listed as shown in Figure 10.

6.4 Experiment-3

As Experiment-3, the query shown in Figure 11 is issued to the system. This query means “Select the images which were painted after 1990, and put them in order, according to the context ‘light’ and ‘bright’.” The images painted after 1990 are selected and ordered, according to the context “light, bright.” The results are shown in Figure 12.

After the selection by the pattern matching, data items are put in order, according to the given context by the semantic associative search. These experiments have clarified the feasibility and the advantage of our associative search method with symbolic filtering and semantic ordering functions.

7 CONCLUSION

In this paper, we have proposed the new associative search method with the symbolic filtering and semantic ordering functions. The symbolic associative search is based on the simple pattern-matching-based function. This function is used as the information filter which repeatedly executes pattern-matching-based comparisons between data items. The semantic associative search function extracts semantically related information by the mathematical semantic operations based on the mathematical model of meaning.

We have also presented the implementation method of the associative search method in the multidatabase environment. This implementation method makes it possible to integrate the existing subsystems for symbolic and semantic associative search under the meta-level system.

As our future work, we will realize a learning mechanism for adapting the information extraction according to individual variation. We will also consider

analytical evaluation of our associative search method. Furthermore, we will study automatic metadata creation(Kashyap et al. 1996) from media data, for designing an advanced multimedia database system.

ID	Title	Author	Year
chagall1	An.Atelier	Chagall	1910
chagall2	A.Grey.House	Chagall	1917
chagall3	A.Gate	Chagall	1917
chagall4	A.Window	Chagall	1917
corot1	Genor.City	Corot	1834
corot2	La.Locheil	Corot	1851
corot3	La.Chelvalla	Corot	1830
corot4	Wood	Corot	1846
gogh1	Dovinie	Gogh	1888
gogh2	Valley	Gogh	1888
hiro1	Cereta	Hiro	1990
hiro2	A.Street	Hiro	1991
hiro3	Concord	Hiro	1992
hiro4	Louvre	Hiro	1990
hiro5	Venice	Hiro	1990
hokusai1	Kanagawa	Hokusai	1829
hokusai2	Tago.no.ura	Hokusai	1829
hokusai3	Misaka	Hokusai	1829
hokusai4	Sekiya	Hokusai	1829
loirand1	The.Harbour	Loirand	1990
loirand2	Near.Broue	Loirand	1990
loirand3	A.Farm	Loirand	1990
loirand4	A.Stone.Pavement	Loirand	1990
nelson1	pursuit	Nelson	1977
nelson2	whaling_ships	Nelson	1980
renoir1	Venise.brouillard	Renoir	1881
renoir2	Garden.at.Cortot	Renoir	1876
renoir3	La.Casbah	Renoir	1881
renoir4	Estaque	Renoir	1882
sarthou1	Sea	Sarthou	1980

Figure 5 The database used in experiments (*famousimages*)

ID	features
chagall1	vivid quiet substance
chagall2	grief terrible sombre
chagall3	sober dynamic motion
chagall4	shine tender calm
corot1	beautiful grand calm
corot2	beautiful delicate calm
corot3	grief sombre sober
corot4	shine beautiful calm
gogh1	merry delight shine
gogh2	grief terrible sombre
hiro1	twilight grand quiet
hiro2	cheer dim quiet
hiro3	beautiful quiet calm
hiro4	fine shine beautiful
hiro5	fine beautiful calm
hokusai1	dynamic strong motion
hokusai2	fight motion calm
hokusai3	delicate calm quiet
hokusai4	vivid motion speed
loirand1	shine grand calm
loirand2	delight shine calm
loirand3	delight grand calm
loirand4	quiet substance material
nelson1	grand dynamic motion
nelson2	twilight calm quiet
renoir1	dim tender quiet
renoir2	delight dim calm
renoir3	loud bustle crowd
renoir4	fine strong quiet
sarthou1	dynamic motion speed

Figure 6 Matadata of images in experiments

```
(project
  (search-mediadata-by-context
    (select famousimages 'Author '== 'Hokusai )
    ID 'rank 'norm space famousimages user
    4 '(power force) )
  '(Rank ID norm)
)
```

Figure 7 The query for Experiment-1

Rank	ID	norm
1	hokusai1	0.291638
2	hokusai2	0.231356
3	hokusai4	0.221350
4	hokusai3	0.135075

Figure 8 The result of Experiment-1

```
(project
  (search-mediadata-by-context
    (select famousimages 'Year '> '1950 )
    ID 'rank 'norm space famousimages user
    30 '(light bright) )
  '(Rank ID norm)
)
```

Figure 9 The query for Experiment-2

Rank	ID	norm
1	hiro4	0.340773
2	loirand1	0.331999
3	loirand2	0.331126
4	hiro1	0.286002
5	nelson2	0.267768
6	loirand4	0.249276
7	hiro3	0.231995
8	sarthou1	0.223989
9	nelson1	0.223900
10	hiro5	0.206538
11	hiro2	0.205272
12	loirand3	0.204344

Figure 10 The result of Experiment-2

```
(project
  (search-mediadata-by-context
    (select famousimages 'Year '>= '1990 )
    ID 'rank 'norm space famousimages user
    30 '(light bright) )
  '(Rank ID norm)
)
```

Figure 11 The query for Experiment-3

Rank	ID	norm
1	hiro4	0.340773
2	loirand1	0.331999
3	loirand2	0.331126
4	hiro1	0.286002
5	loirand4	0.249276
6	hiro3	0.231995
7	hiro5	0.206538
8	hiro2	0.205272
9	loirand3	0.204344

Figure 12 The result of Experiment-3

REFERENCES

- Bright, M.W., Hurson, A.R., and Pakzad, S.H. (1992) A Taxonomy and Current Issues in Multidatabase System, *IEEE Computer*, Vol.25, No.3, pp.50-59.
- David, R., and Lenat, D.B. (1982) *Knowledge-based systems in artificial intelligence*, McGraw-Hill Book Co.
- Deerwester, S., Dumais, S. T., Landauer, T. K., Furnas, G. W. and Harshman, R. A. (1990) Indexing by latent semantic analysis, *Journal of the Society for Information Science*, vol.41, no.6, pp.391-407.
- Kashyap, V., Shah, K., Sheth, A. (1996) Metadata for Building the Multimedia Patch Quilt, V.S. Subrahmanian, S. Jajodia, eds., *Multimedia Database Systems*, pp.297-319.
- Kitagawa, T. and Kiyoki, Y. (1993) The mathematical model of meaning and its application to multidatabase systems, *Proceedings of 3rd IEEE International Workshop on Research Issues on Data Engineering: Interoperability in Multidatabase Systems*, pp.130-135.
- Kiyoki, Y., Kitagawa, T. and Hayama, T. (1994) A metadatabase system for semantic image search by a mathematical model of meaning, *ACM SIGMOD Record*, vol. 23, no. 4, pp.34-41.
- Kiyoki, Y., Kitagawa, T. and Hitomi, Y. (1995) A fundamental framework for realizing semantic interoperability in a multidatabase environment, *Journal of Integrated Computer-Aided Engineering*, Vol.2, No.1, pp.3-20, John Wiley & Sons.
- Kolodner, J.L. (1984) *Retrieval and organizational strategies in conceptual memory: a computer model*, Lawrence Erlbaum Associates.
- Krikelis, A., Weems C.C., (1994) Associative processing and processors, *IEEE Computer*, Vol.27, No. 11, pp.12-17.
- Litwin, W., Mark, L., and Roussopoulos, N., (1990) Interoperability of Multiple Autonomous Databases, *ACM Comp. Surveys*, Vol.22, No.3, pp.267-293.
- Longman Dictionary of Contemporary English, Longman.
- Comm. ACM (1996) *Natural language processing*, Comm. ACM, Vol.39, No.1.
- Ogden, C.K. (1940) *The General Basic English Dictionary*, Evans Brothers Limited.
- Potter J.L. (1992) *Associative Computing*, *Frontiers of Computer Science Series*, Plenum.
- Raju, K. V. S. V. N. and Majumdar, A.K. (1988) Fuzzy Functional Dependencies and Lossless Join Decomposition of Fuzzy Relational Database Systems, *ACM Transactions on Database Systems*, vol.13, no.2, pp.129-166.
- Rundensteiner, E.A., Hawkes, L.W. and Bandler, W. (1989) On Nearness Measures in Fuzzy Relational Data Models, *International Journal of Approximate Reasoning*, vol.3, no.3, pp.267-298.

Sheth, A.P., and Larson, J. (1990) Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases, *ACM Computing Surveys*, Vol.22, No.3, pp.183-236.

UniSQL (1995) UniSQL/X User's Manual, UniSQL, Version 3.1.

8 BIOGRAPHY

Naofumi Yoshida was born in 1972. He is currently a master course student of University of Tsukuba. He is engaged in the research of database systems.

Yasushi Kiyoki received his BS, MS and PhD degrees in electrical engineering from Keio University in 1978, 1980 and 1983, respectively. He is currently an associate professor of Department of Environmental Information at Keio University. During 1990 and 1991, he was a visiting researcher in University of California at Irvine and University of Texas at Austin. His research addresses multidatabase systems, semantic associative processing, and multimedia database systems.

Takashi Kitagawa was born 1956 in Gifu, Japan. He received his PhD from Nagoya University, Doctoral course in Computer Science. He was a visiting scholar at Dept. of Computer Science, Stanford University, 1980-81, Assistant Professor at Dept. of Mathematics, Ehime University, 1982-89. He is currently an Associate Professor at Inst. of Information Sciences and Electronics, University of Tsukuba. He is engaged in inverse problems, numerical analysis, multimedia information systems.