Paper

Similarity of Compounds and Reactions Based on Self Organized Conceptual Structures of Organic Synthesis Information

Jianghong AN and Yuzuru FUJIWARA

Institute of Electronics and Information Science, University of Tsukuba
Tsukuba-shi, Ibaraki 305, Japan

A data model and a information base system for organic syntheses are introduced. Self organization of information and generation of virtual concepts and candidate solutions which are essential for open world relevant systems are described. The original information of the system includes comprehensive typical reactions and related information of huge quantity. Conceptual and logical structures of information are constructed by self organization of semantic relationships in original information. The system provides functions with semantic processing such as analogical reasoning and abduction because similarity of compounds and reactions can be measured by using conceptual structures.
Keywords: similarity, analogical reasoning, information base system, self organization, conceptual structures, organic synthesis, semantic processing, thinking support system.

1. Introduction

The information for organic syntheses consists of the information about compounds and reactions. This information is used not only for retrieval of known reactions and compounds but also for prediction of new reactions and for design of new organic syntheses.

It is said that the number of known compounds is over 10 millions. These compounds and the reactive relations among them constitute a huge information space and computer-based information systems are indispensable because the management and utilization of the information is far beyond feasibility of manual methods.

In the field of organic syntheses, a trial and error method is often used based on knowledge of compounds (which was acquired from research on analytical, physical and structural chemistry) and reactions. A lot of computer-based systems were developed to support chemists.[1)2)3)] They can be divided into two types according to the information/knowledge stored in systems: one is an expert system which stores reaction rules and the other one is a database system which stores reaction examples.

However, the acquisition and organization of reaction rules (knowledge of reactions) are very difficult for expert systems because of the diversity of compounds and reactions. On the other hand, database systems can not deal with unknown compounds and reactions although the purpose of synthesis is to get new compounds and design new reaction processes.

Approaches such as analogical reasoning and case-based reasoning are to solve these problems. Analogical reasoning transforms facts and relationships among objects in a base domain into a target domain according to the similarity between the objects of the base and the target domains. However, it is assumed that the base of analogy can be chosen in most cases, *i.e.* a very popular example is about the analogical reasoning of the structure of atoms from the solar system. The discussion is focused on the similarity between the atom and the solar system, but why and how the solar system could be taken as a candidate were not explained.[4)5)6)7)] Moreover, little attention has been given to mechanisms for searching the huge space of practical information instead of dealing with small information space. On the other hand, case-based reasoning takes advantage of similar cases accumulated in systems to transform solutions directly. Several systems were reported so far but methods for organizing and retrieving relevant case bases were not mentioned because the number of cases is very small, *e.g.* the order is from 10 to 1000.[8)9)]

The goal of this work is to develop a novel thinking-support information system (TSI system) for domain experts. The main points are shown bellow:

1) Collection of comprehensive reaction examples and related information of huge quantity, which are essential to the TSI system. In order to support experts, the system should possess more knowledge or organized information than experts do.

2) Construction of information spaces by self organization: the accumulated information must be organized according to semantic relationships. Furthermore, the self organization method should be applied because of great volume of information and evolution of systems.

3) Manipulation of the organized information for semantic processing such as analogical reasoning: In order to support thinking activity of experts, semantic processing is indispensable.

The section 2 describes the method of self organization of organic synthesis information structures. In the section 3 the measurement of similarity based on conceptual structures is given. The section 4 illustrates analogical reasoning of organic reactions by an example. The discussion and conclusion are given in the section 5 and the section 6 respectively.

2. Self organization of organic synthesis information structures

The term "information structures" is used to refer to the structures which represent relationships among concepts, such as conceptual hierarchy, causal ones and so on.[10)11)12)13)14)]

The following three kinds of structures corresponding to the semantics of information are realized:

1) Conceptual structure: The structure which represents relationships such as equivalence, inclusion, is-a, part-of relation, and so on.

2) Logical structure: The structure which represents relationships such as causal relation, logical order, dependency, and so on.

3) Physical structure: The structure which represents the internal storage and locations of source information.

The process of organizing information is that of machine learning. The self organization approach is devised here. The term "self organization" is used in the sense that the structures are constructed based on the semantic relationships which are automatically extracted from original information by the system. The process may be carried out repeatedly in order to refine the constructed structures. The self organization is also necessary for evolution of the structures.

2.1 Information of organic syntheses

The organic synthesis information is stored in a data base of organic reaction named CORES.[15)] The reaction information is composed of starting compounds, products and reaction conditions. It can be expressed as following:

$$X_1 + X_2 + \cdots + X_m \xrightarrow{C} Y_1 + Y_2 + \cdots + Y_n,$$

where X_i (i=1,m) is a starting compound, Y_i (i=1,n) is a product and C is a reaction condition. Compounds are represented in structural formula with connection tables. Reaction conditions include detailed information such as reagents, solvents, catalysts, temperature, pressure, yield and so on.

About 50,000 reactions have been collected and stored as the source of information of the system.

2.2 Construction of logical structures of reactions

The network of reactions generated based on the collection of reactions is called logical structures of reactions. This network can be generated automatically by piling up common compounds for reactions. A graph isomorphism algorithm is necessary to decide if two connection tables represent a same compound.

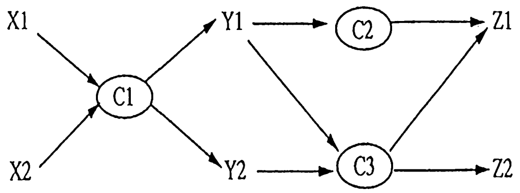
For example, from a set of the following three reactions:

$$X_1 + X_2 \xrightarrow{C_1} Y_1 + Y_2$$

$$Y_1 \xrightarrow{C_2} Z_1$$

$$Y_1 + Y_2 \xrightarrow{C_3} Z_1 + Z_2$$

the following network can be generated.



Considering that the starting compounds and products are not necessarily single, a hypergraph [21] or more flexible representation given below is necessary.[13)14)]

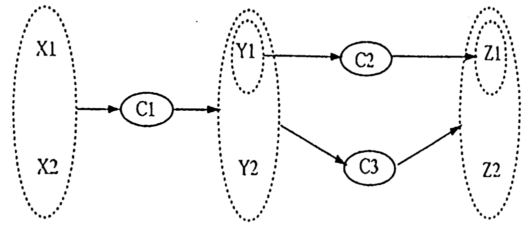


Fig. 1 shows a part of the logical structure of the system. The numbers in the ellipses stand for the identification codes of reactions.

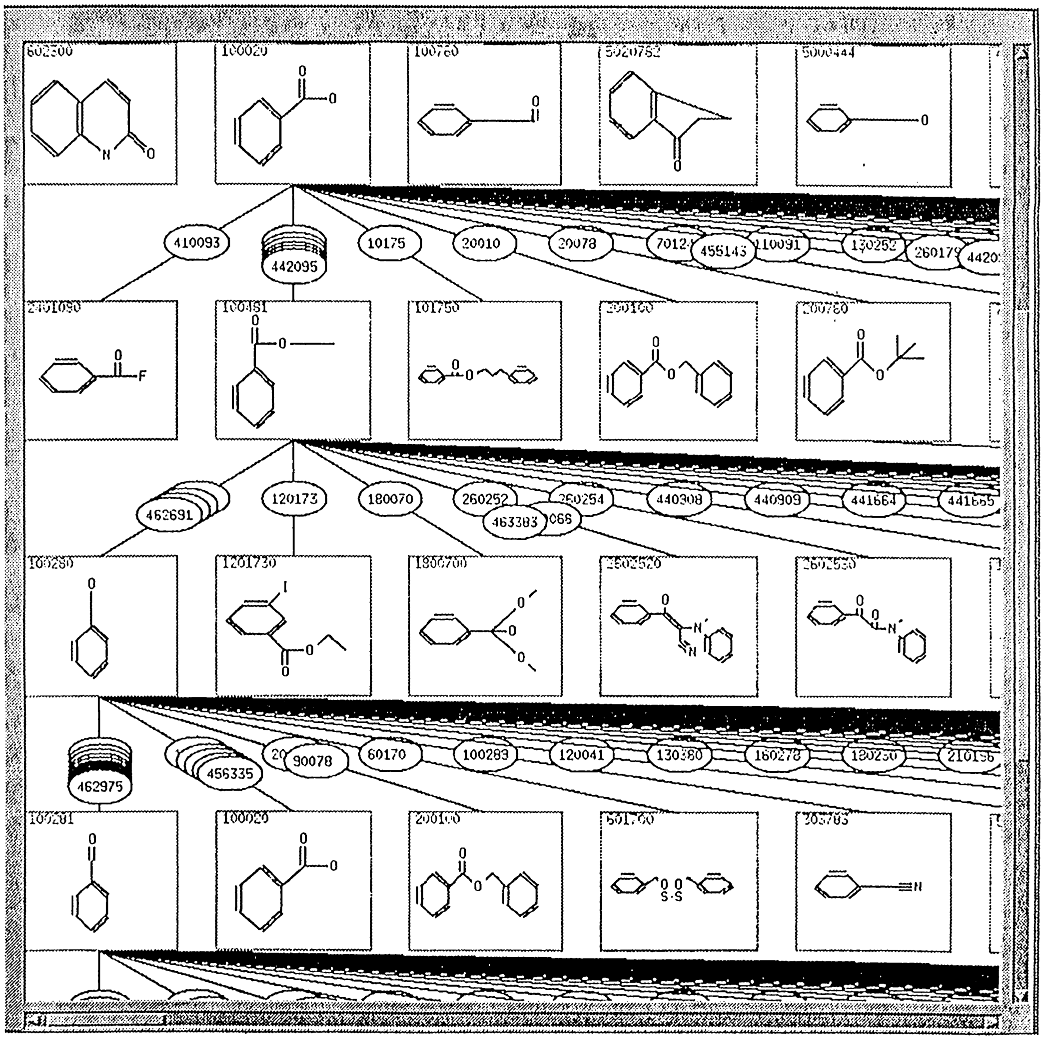
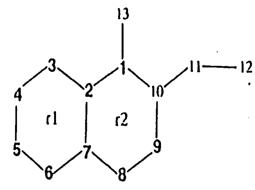


Figure 1. A part of a logical structure

2.3 Construction of conceptual structures of compounds

The conceptual structure of compounds is constructed as a conceptual hierarchy based on the inclusive relation between chemical structures of compounds because properties of compounds mainly depend on structures of molecules.

The basic algorithm can be described as follows: from each structure in the database, substructures are generated by removing bonds, atoms or super-atoms (ring) and tested recursively and the hierarchy of structures is organized.



LN = { 12, 13 }

SRE(r1) = { (2, 3), (3, 4), (4, 5), (5, 6), (6, 7) }

SRE(r2) = { (1, 2), (7, 8), (8, 9), (9, 10), (1, 10) }

SRN(r1) = { 3, 4, 5, 6 }

SRN(r2) = { 8, 9 }

|SRE(r1)| - |SRN(r1)| = 5 - 4 = 1, r1 is a TR.

|SRE(r2)| - |SRN(r2)| = 5 - 2 = 3, r2 is not a TR.

Figure 2. Example of molecule structure and notations

The following notations are introduced to make the description clear:

Algorithm

```
Last := the last number of the structure in the database;
Current := 1;  /*  the current structure to be processed  */
WHILE ( Current <= Last) DO {
   IF ( Current is a primitive structure) THEN
             /* atoms or basic structures which have proper names */
             /* in nomenclatures of compounds                     */
       Current := Current + 1;   /* skip the structure */
   ELSE
      detect the SSSR(the Smallest Set of Smallest Rings) of Current;
      FOR ( each LN and TR of Current ) DO {
         remove the LN or TR to generate subgraph SG;
                     /* generate the substructures */
         CALL ISOM subroutine to determine if SG exists in the DB;
                     /* search the database  */
         IF ( not existence ) THEN
            Last := Last + 1;
            id of SG := Last;
            register SG in the Database;
         END-IF
         c_id :=  id of SG;
         add c_id to the boarder concept list of Current;
         add Current to the narrower concept list of c_id;
                /* generate conceptual structure  */
      } /* end of FOR */
      Current := Current + 1;   /* next structure */
   END-IF
} /* end of WHILE */
```

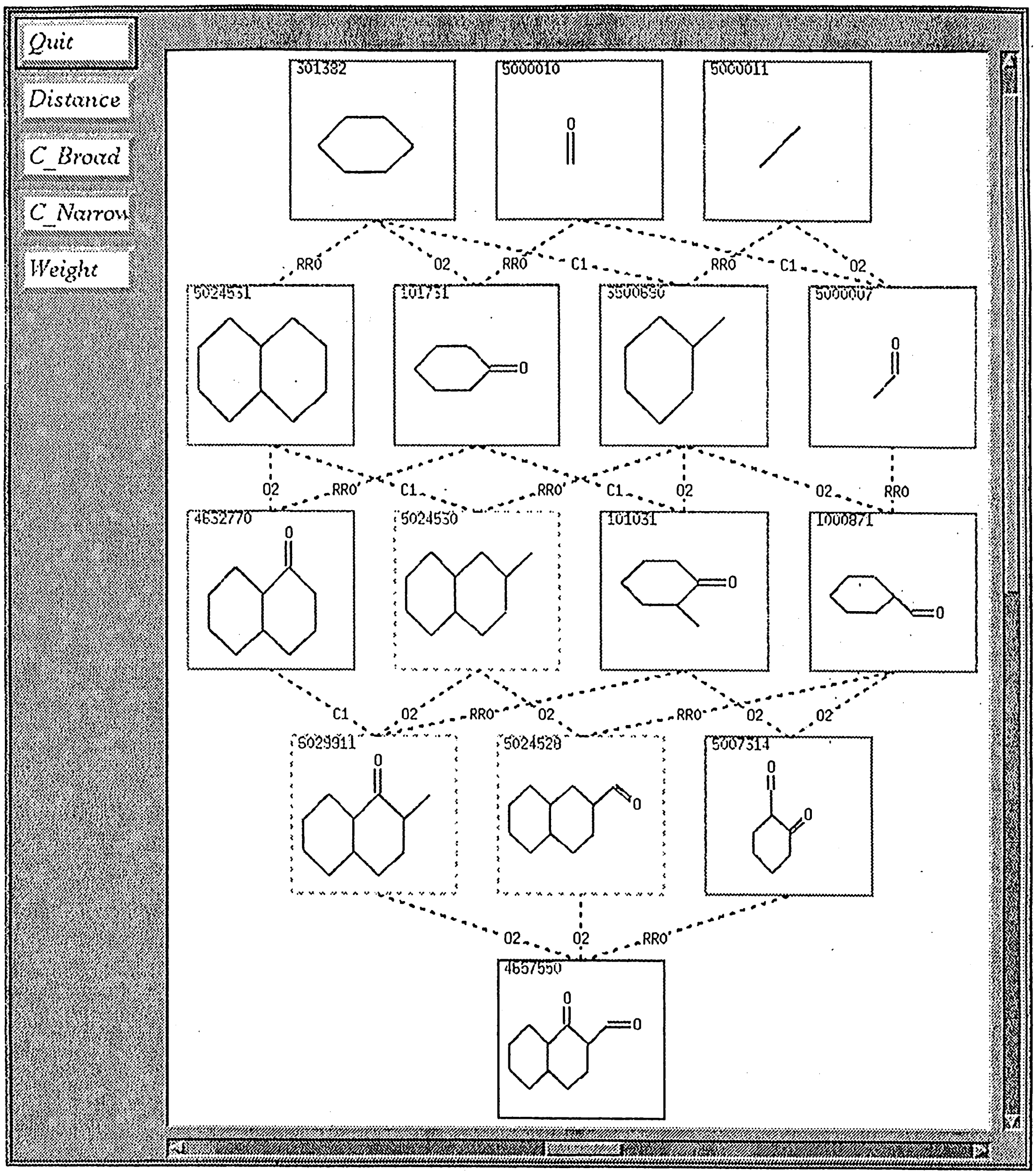


Figure 3. Process of organizing conceptual structure by generating method

Degree of a node *n*, *D*(*n*): The number of adjacent nodes of node *n*.

Leaf Node, *LN*: node *n* with *D*(*n*) = 1.

Simple Ring Edge, *SRE*: an edge which belongs to one and only one ring. *SRE*(*r*) is the set of simple ring edges in ring *r*.

Simple Ring Node, *SRN*: a node *n* which belongs to a ring and *D*(*n*)=2. *SRN*(*r*) is the set of simple ring nodes in a ring *r*.

Terminal Ring, *TR*: a ring *r* with |*SRE*(*r*)| – |*SRN*(*r*)| = 1.

Figure 2 illustrates an example of molecule structure and notations used in the algorithm.

In order to make the description of the algorithm simple, suppose the structures are numbered from 1 to the total number of the structures. Moreover, some detailed processes are omitted.

In the algorithm, the substructures generated are connected graphs because only leaf nodes and terminal rings are removed from structures. The conceptual structure of compounds is self organized systematically by a generating -- searching cycle.

The subroutine used to detect the rings is based on the *SSSR* algorithm.[16] It finds the smallest set of smallest rings in a structure. The graph isomorphic determination subroutine *ISOM* is a improved version of the E. H. Sussenguth's algorithm.[17]

Fig. 3 shows the process of generating substructures and constructing the hierarchy. Fig. 4 shows a part of the conceptual structure of the system. The structure is composed of nodes which stand for compounds and links which stand for inclusive relationships between nodes. The link has a label which describes difference between a pair of nodes. Moreover, the nodes which have solid borders stand for compounds which are stored in the database and dash borders stand for virtual compounds generated by the system.

Fig.1, Fig.3 and Fig.4 show clearly that the conceptual and logical structures have to be constructed in a self organizing manner because of their complexity and huge size.

2.4 Construction of conceptual structures of reactions

The conceptual structure of reactions represents is-a relationships among reactions. Information of organic reaction consists of starting compounds, products and reaction conditions. Reaction conditions include detailed information such as reagents, solvents, catalysts, temperature, pressure, yield and so on.

The conceptual structure of reactions is constructed according to generality of change of chemical structures from the starting compounds to the products together with the reaction conditions.

The approach of constructing the structure is a bottom-up manner, that is, using more generic changes on chemical structures to organize reactions, then taking advantage of the knowledge on reaction patterns such as oxidation, reduction, replacement, and so on.

Fig. 5 illustrates a example of the conceptual structure of reactions. The numbers in ellipses stand for the identification codes of reactions.

2.5 Application of the conceptual and logical structures

The conceptual and logical structures generated above can be used for many purposes.

2.5.1 Conventional application

Compounds searching:

1) Searching by substructures: searching compounds which include query substructures;

2)Searching by super structures: searching compounds which are included in query structures;

3) Searching similar structures: searching compounds which are near to a query structure in the conceptual or logical structures:

3.A) Compounds with common substructures;

3.B) Compounds with similar reactive properties:

3.B.a) Starting compounds (or products) in which similar functional groups react in similar conditions;

3.B.b) Starting compounds which lead to similar products in similar conditions;

3.B.c) Products which are produced from similar starting compounds in similar conditions;

3.B.d) Compounds which are near to each other in the logical structure.

Reaction searching:

1) Searching reactions by structures of compounds;

2) Searching synthesis routes to a query compound;

3) Searching similar reactions by structure of compounds.

Browsing and navigation:

1) Browsing the whole or a part of conceptual and logical structures;

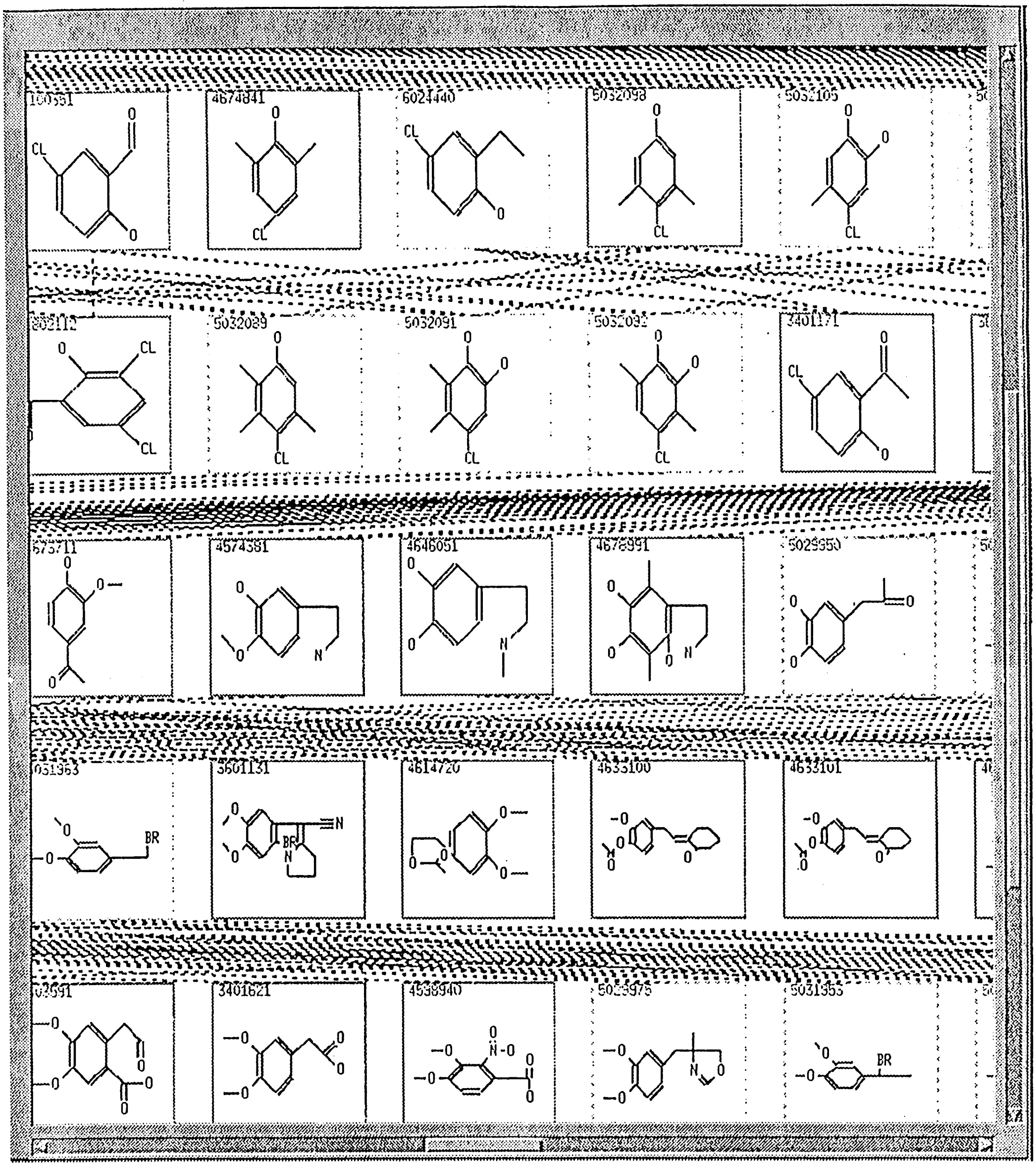


Figure 4. A part of a conceptual structure of compounds

2) Navigation in conceptual or logical structures;

3) Switch over from conceptual structures to logical structures or in the opposite direction.

2.5.2 Advanced application

The functions listed above are useful but not sufficient for synthesis chemists because the ultimate aim of organic synthesis is to synthesize new compounds or to design new synthetic routes. It can not be satisfied simply by searching or browsing.

The following knowledge is essential for chemists:

1)Knowledge on structures and properties of compounds;

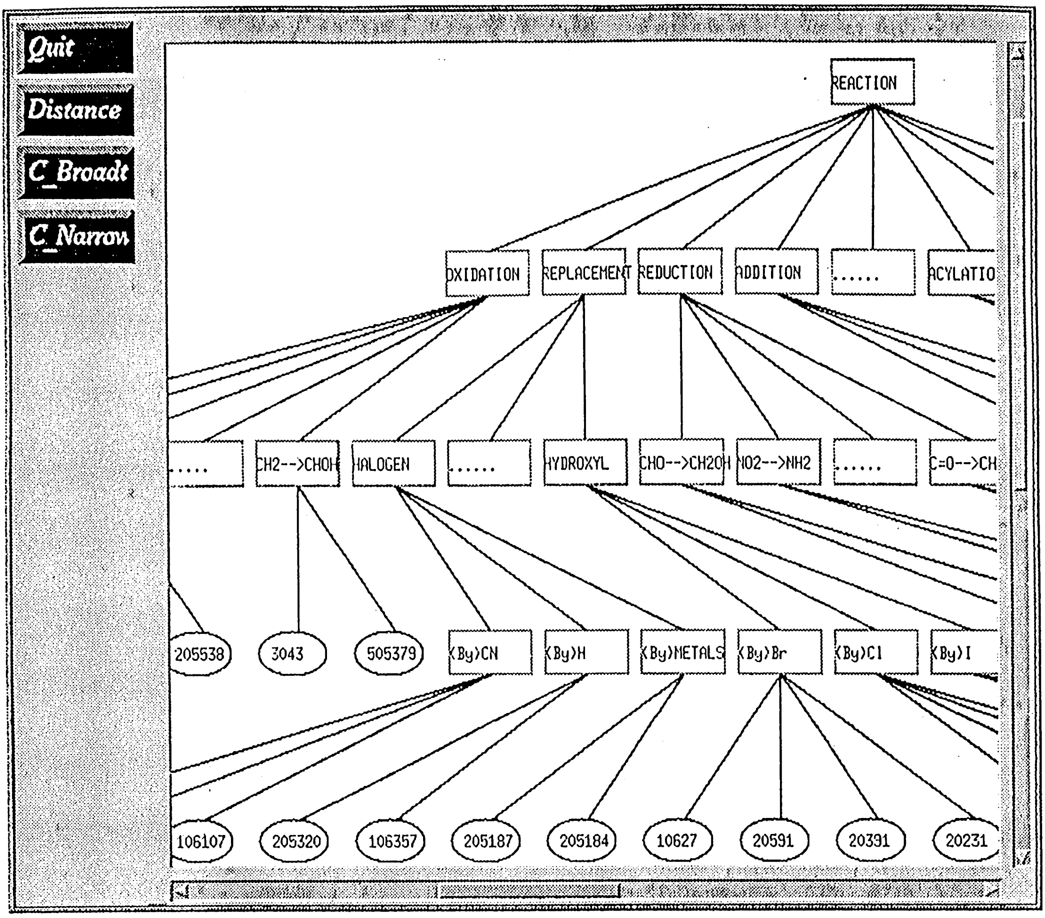


Figure 5. A part of the conceptual structure of reactions

2) Knowledge on reactions;

3) Sufficient examples of reactions.

It is clear that computer systems can store and deal with far more examples of reactions than individual experts. The key point is the ability of analyzing examples and solving new problems. The ability depends on how much knowledge of compounds and reactions the system possessed. The purpose of organizing information in the conceptual structures and the logical ones is to acquire and understand the knowledge automatically and represent it in a form of easy access. In this system, the conceptual structures of compounds correspond to the knowledge of compounds and the conceptual structures of reactions correspond to the knowledge of reactions.

It is essential for synthesizing new compounds that the systems can deal with unknown compounds, that is to say, the systems should be able to predict the properties of new compounds. This is the reason why the system generates virtual compounds in the process of constructing the conceptual structures of compounds. The properties of these virtual compounds can be estimated based on the compounds close to it in the conceptual structures. The logical structures should also be augmented to generate the candidates of solutions for new reactions.

However, the augmentation must be restricted because the information space of logical structures may be expanded infinitely like conceptual structures. The restrictions are listed below:

1) To extend when a query of new reaction is carried out;

2) To extend to reactions which are sufficiently reasonable.

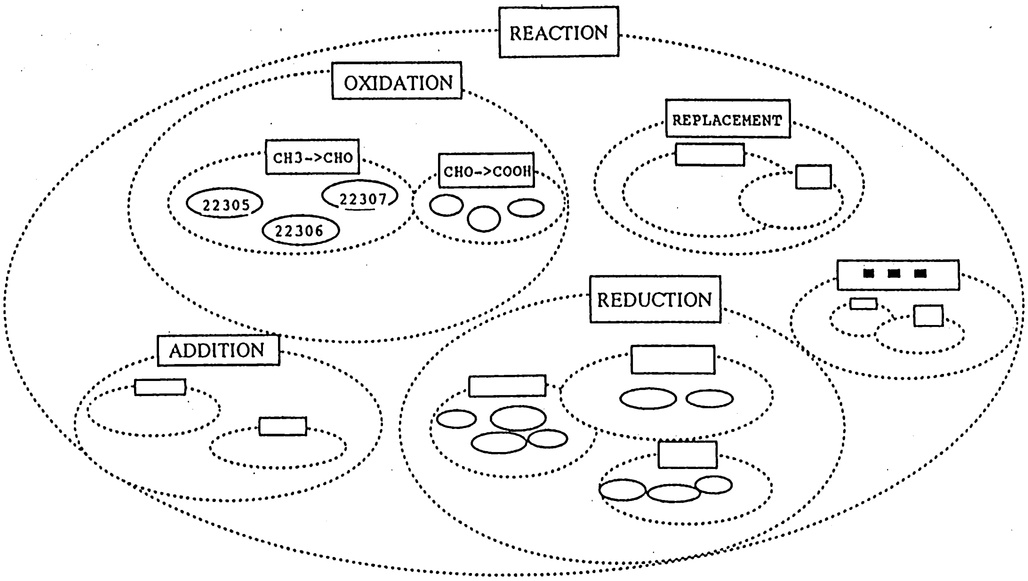


Figure 6. A part of conceptual structure of reactions

Analogical reasoning is useful for extending the logical structures. Moreover, the degree of confidence and the grounds of the reasoning should be kept in the extended reactions.

3. Measurement of similarity by conceptual structures

3.1 The model of organic synthesis information

Let A be a set of atoms which includes generic ones and pseudo ones beside normal ones.

$$\kappa = < A, E, B >$$

makes up the space of chemical structures,

where $E \subseteq A \times 2^A$ is a set of connections between atoms, and
B is a set of types of connections, *i.e.* bonds, which includes generic ones also.

Notice that the one to many relationship $A \times 2^A$ is introduced for coordination (complex) compounds.

The information space of organic synthesis is constructed based on κ.

1) The conceptual structure of compounds:

$$\Gamma_1 = < \kappa, \Omega_1, L_1 >,$$

where $\Omega_1 \subseteq \kappa \times 2^\kappa$ stands for the hierarchical relationships between compounds, and
L_1 is a set of labels for the differential relationships.

2) The logical structure of reactions:

$$\Gamma_2 = < \kappa, \Omega_2, L_2 >,$$

where $\Omega_2 \subseteq 2^\kappa \times 2^\kappa$ stands for the reactive relationships, and
L_2 is a set of labels for the reaction rules and conditions.

Notice that generic reactions, *e.g.* reaction patterns are included in the structures.

3) The conceptual structure of reactions:

$$\Gamma_3 = < \Gamma_2, \Omega_3, L_3 >,$$

where $\Omega_3 \subseteq \Gamma_2 \times 2^{\Gamma_2}$ stands for the hierarchical relationships between reactions, and
L_3 is a set of labels for the differential relationships including hierarchical ones.

The structure is an extended hypergraph.[13)14)21)] Therefore, Fig. 5 can be represented in Fig. 6 equivalently.

The information space of organic syntheses is defined as following:

$$\Gamma = \Gamma_1 \cup \Gamma_2 \cup \Gamma_3 \cup \Delta$$

Γ is a hybrid model with more than one kind of nodes, links and labels. Δ stands for the information structures which are not described here such as the conceptual structure of reagents. The components of Γ are connected each other in various ways.

3.2 The measurement of similarity

The similarity is a popular relationship because any two objects may be similar in a sense. However, the meaning of similarity is dependent on viewpoints. The relationship of similarity is defined as partial overlap between the concepts. Therefore, it is natural that the detection and measurement of the relationship of similarity should be carried out by detecting and measuring overlap between concepts. The conceptual structures can be used to deal with similarity because overlaps between concepts are represented by the structures.

Compounds are taken as an example to show the measurement of similarity of concepts. Compounds are represented as bellow:

$$C = (C_{st}, C_{pr}),$$

where $C_{st} = (C_{st1}, C_{st2}, \cdots, C_{stm})$ is a set of functional groups or substructures as the descriptors of structures of compounds, and

$C_{pr} = (C_{pr1}, C_{pr2}, \cdots, C_{prn})$ is a set of physical, chemical and other properties of compounds as the descriptors of compounds.

The distance between a couple of compounds can be computed by the inner product of the vector corresponding to the difference in compounds and the weight vector , that is :

$$Distance(C_i, C_j) = W \cdot | C_i - C_j |,$$

where W is a weight vector corresponding to difference of a pair of compounds, which reflects viewpoints of similarity. W is a function of reaction R, that is, $W = f(R)$ in cases of syntheses.

Reactions are represented as below:

$$R = (Reaction\ pattern, Reagent, Solvent, Catalyst, Temperature, Pressure, Yield, \cdots).$$

It means that similarity of compounds varies with the reaction pattern, reagent, etc. of the reaction concerned. It is different from other researches in which only static similarity is measured.[18)]

Knowledge on organic syntheses is required to decide the function f . This kind of knowledge can be acquired from systematic knowledge of organic chemicals in text books and from examples of reactions by way of statistics, induction, and so on.

The difference in structures and properties of

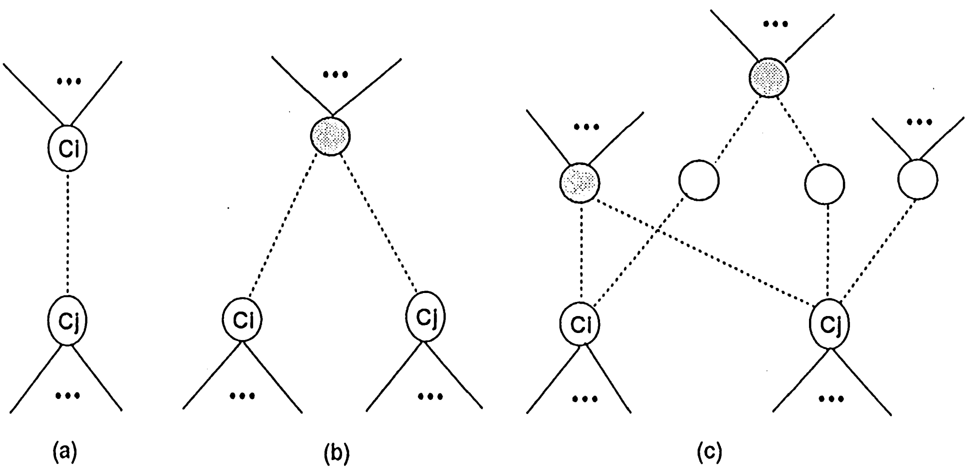


Figure 7. Paths between compounds in conceptual structure

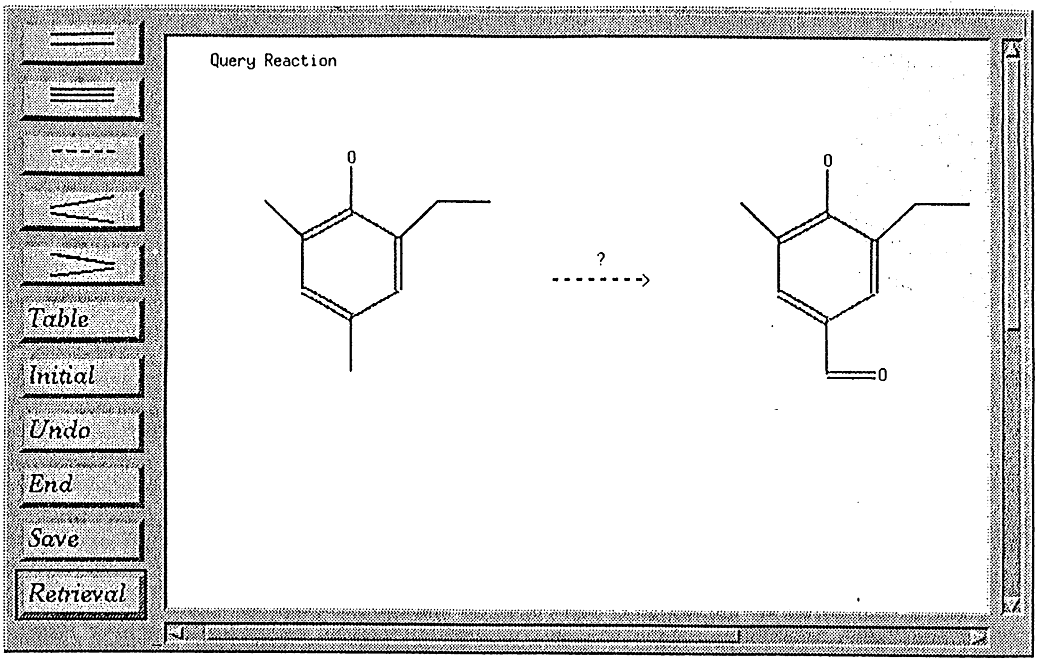


Figure 8. An example of query reaction

compounds is represented by conceptual structures in the system. Compounds with common substructures are connected by paths in conceptual structures. The difference can be calculated based on the labels corresponding to the paths. The algorithm for calculating the difference is to collect the labels corresponding to the shortest paths between compounds, because the difference is not dependent on the paths.

Figure 7 shows cases of paths between compounds C_i and C_j. (a) is the case of that C_i is a substructure of C_j, (b) is the case of that C_i and C_j with only one common substructure and (c) is the case of that C_i and C_j with more than one common substructures.

The problem of detecting common substructures of a pair of compounds is NP complete in graph theory.[20] Therefore, except the case of that one or both compounds are very small molecules, detecting common substructures using conceptual structures is efficient than that using graph matching algorithms directly.

The similarity of concepts can be measured based on the distance between them.

4. Analogical reasoning of reactions

Analogical reasoning is one of the powerful tool for problem solving and machine learning. In our systems, analogical reasoning is used to infer whether a reaction can take place between a pair of sets of compounds, and the reaction conditions if it can take place. The ability of inference is necessary not only for intelligent retrieval of reactions but also for extending the logical structure of reactions. Fig. 8 shows an example of query reaction which is not in the database. Therefore, analogical reasoning is to be used to give an answer.

The procedure of analogical reasoning of reactions can be described as follows:

1) To recognize a reaction pattern by detecting change of structures between starting compounds and products. In the example, the reaction pattern is $CH_3 \rightarrow CHO$.

2) To retrieve candidate reactions which share the same pattern with the query reaction. Taking advantage of the conceptual structures of

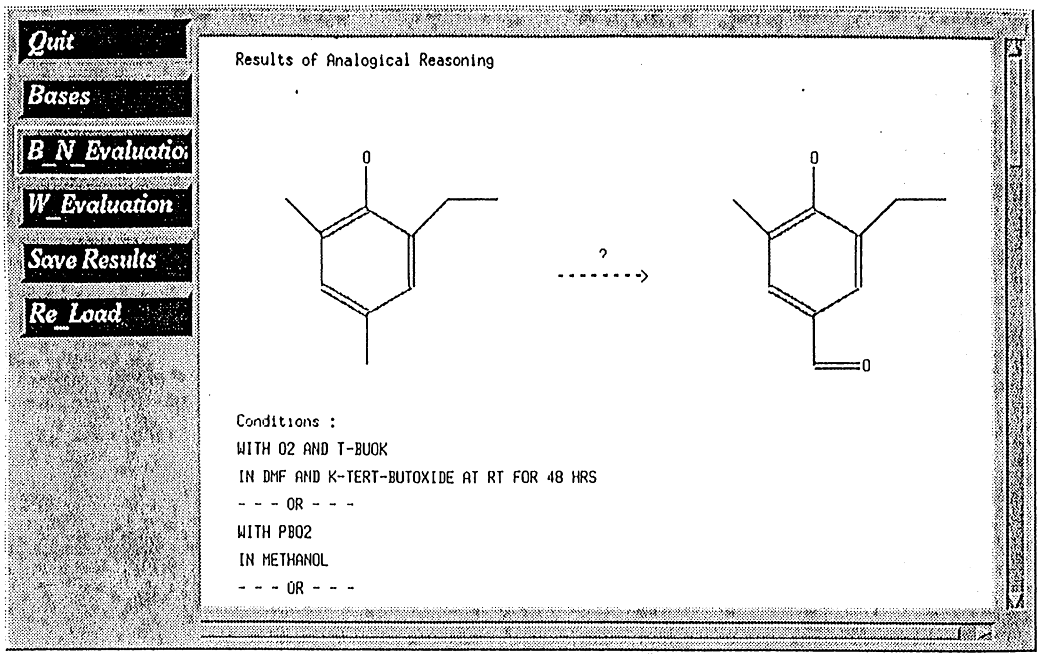


Figure 9. The result of analogical reasoning

reactions, the reactions with same pattern can be accessed efficiently.

3) To measure similarities between candidates and the query reaction. The similarities depend on the relevant compounds and the reaction conditions between them. The conceptual structure of compounds can be used to measure the similarities of compounds.

4) To select a set of similar reactions from the candidates according to similarity and threshold;

5) To apply reaction conditions of similar reactions to the query reaction: In this step, comparisons are made not only between the query reaction and similar reactions but also among the similar reactions.

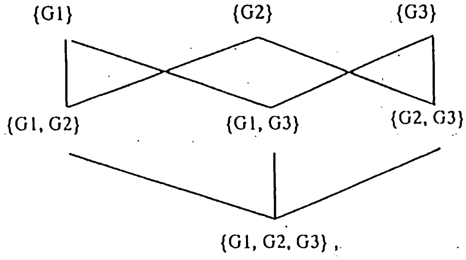
Fig. 9 illustrates the result of the example of analogical reasoning. The list shows a set of applicable reaction conditions for the query reaction.

Fig. 10 and Fig. 11 show two examples of the analogies (similar reaction) of the query reaction. The list of similar reactions can be displayed by pushing the "Bases" button in the result display screen (Fig.9).

5. Discussion

A large number of researches on classification and structure searching of compounds based on functional groups have been performed[18)19)20)] because properties of compounds mainly depend on functional groups included in the structures. However, it is difficult to choose a particular set of functional groups used for classification and access because the set is dependent on viewpoints of application. Using fixed set is not suitable for investigating the properties of compounds from various viewpoints.

The conceptual structure constructed here can be used to solve this problem. For example, let's suppose that a set of functional groups is $FG = \{G_1, G_2, G_3\}$ according to the current viewpoint. The conceptual structure corresponding to FG is a lattice as follows:



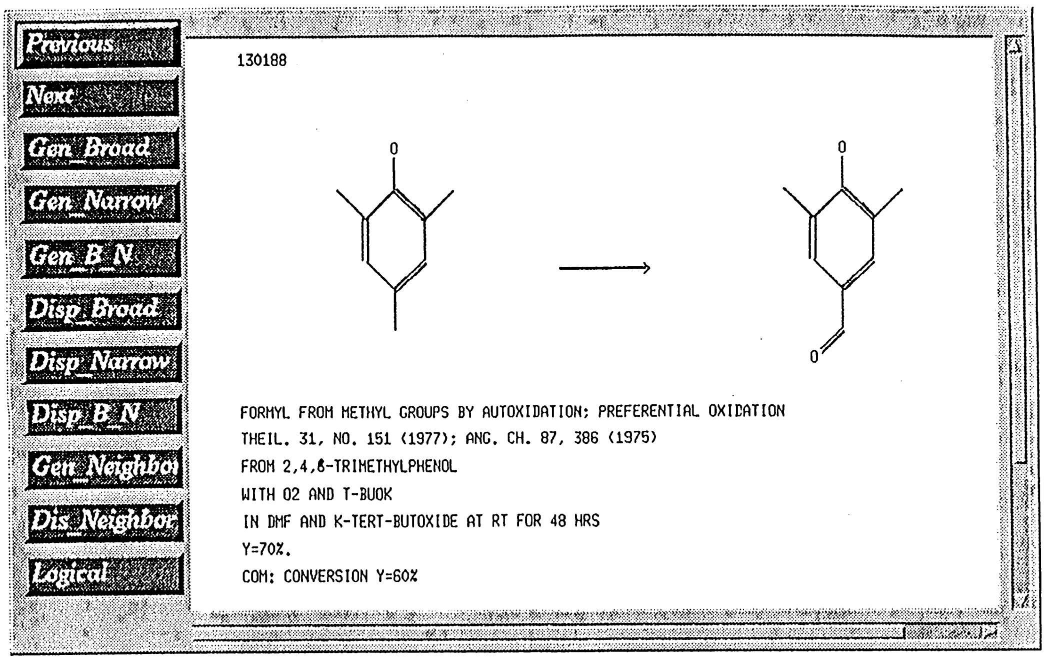


Figure 10. An analogy of the query reaction

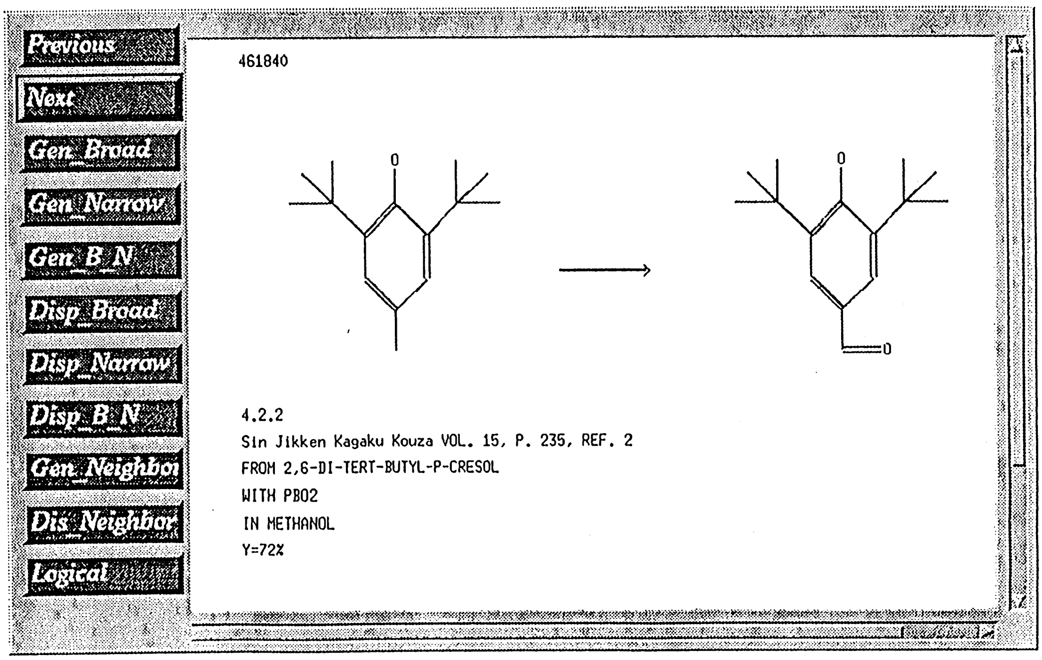


Figure 11. Another analogy of the query reaction

where $\{G_i\}$ is a set of structures that includes functional groups $\{G_i\}$, $\{G_i, G_j\}$ is a set of structures that includes G_i and G_j, $\{G_i, G_j, G_k\}$ is a set of structures that includes G_i, G_j, and G_k.

The lattice can be generated as follows because G_i (i=1, 3) corresponds to nodes in the conceptual structure of compounds of the system.

$\{G_i\}$ = *Narrow*(G_i)
$\{G_i, G_j\}$ = *Narrow*(G_i) O *Narrow*(G_j)
$\{G_1, G_2, G_3\}$ = *Narrow*(G_1) O *Narrow*(G_2) O *Narrow*(G_3),

where *Narrow*(G_i) stands for the set of narrower nodes including G_i in the conceptual structure of compounds.

This example shows that conceptual structure reflecting temporal viewpoints can be generated for any set of functional groups because conceptual structures of the system contains every substructure of the compounds as its nodes.

By organizing information structures, it become possible to take a broad view of the information space (concepts and the relationships among them), that is, see things in perspective.

In analogical reasoning and case-based reasoning approaches, selection of the analogies and measurement of the similarities are not easy tasks because not only the space of analogies is very large but also the measurement of similarities is relevant to the problem of semantic processing. Moreover, the evaluation of the results of reasoning is the most difficult problem for the conventional systems. Organized information structures such as conceptual structure and logical structure can play an important role in solving these kinds of problems because the relationships among facts or concepts are represented effectively.

Taking advantage of the structured information, it is possible to generate lacking or new information, for example, to generate some new concepts or to establish some connections among concepts. This kind of functions are essential to the thinking support systems.

6. Conclusion

The capability of thinking support systems mainly depends on semantic processing of comprehensive information. In order to deal with meaning of information effectively the information must be organized in conceptual structures to represent diversified semantic relationships. Moreover, the conceptual structures have to be self-organized because they are both complicated and huge in quantity beyond manual construction. The method and system for self organizing organic synthesis information are described. The structures of information consist of logical structures of reactions, conceptual structures of compounds and conceptual structures of reactions. The measurement of similarity and analogical reasoning based on conceptual structures are presented. However, some factors involving similarity such as the chemical properties and physical properties can not be dealt with directly based on the conceptual structures. The inclusion of more abstract conceptual structures remains as a future work.

On the other hand, generation of information is essential for thinking support system because the research activities are of open world problems. In the field of organic syntheses, the object is to synthesize new compounds. The generation of virtual compounds and candidate reactions are illustrated. In fact, the generation of compounds plays an important role in constructing the conceptual structures systematically.

References

1) E.J. Corey, W.T. Wipke: "Computer-assisted Design of Complex Organic Syntheses," *Science*, Vol.166, 1969, p.178-192.
2) E.J. Corey, A.K. Long, S.D. Rubinstein: "Computer-assisted Analysis in Organic Synthesis," *Science*, Vol.228, 1985, p.408-418.
3) J. Gasteiger, W. Ihelenfeld, R. Fick, and J.R. Rose: "Similarity concepts for the planning of organic reactions and syntheses," *J. Chem. Inf. Comput. Sci.*, Vol.32, No.6, 1992, p.700-712.
4) Patric H. Winston: "Learning and Reasoning by Analogy," *Communications of the ACM*, Vol.23, No.12, 1980, p.689-703.
5) D. Gentner: "Structure-Mapping: Theoretical Framework for Analogy," *Cognitive Science*, Vol. 7, No.2, 1983, p.155-170.
6) S. Kedar-Cabelli: "Purpose-Directed Analogy," *Proc. of 7th Cognitive Science Society*, 1985, p.150-159.
7) J.G. Carbonell: "Derivational Analogy: A Theory of Reconstructive Problem Solving and Expertise Acquisition," *Machine Learning* II, Morgan Kaufmann, 1986, p.371-392.
8) J.L. Kolodner, R.L. Simpson and K. Sycara: "A Process Model of Case-Based Reasoning in Problem Solving," *IJCAI-85*, 1985, p.284-290.
9) K.J. Hammond: "Case-based Planning: A Framework for Planning from Experience," *Cognitive Science*, Vol. 14, 1990, p.386-443.

10) J.F. Sowa: "Conceptual Structures," Addison-Wesley, 1984.
11) Z.Q. Wang, Y. Fujiwara et al: "Learning and Reasoning in the Information-Base System for Organic Synthesis Research," *Journal of Japan Society of Information and Knowledge*, Vol.2, No.1, 1991, p.71-82.
12) Y. Fujiwara, Z.Q. Wang et al: "The Multicategorical Structures of Information for Inferences and Reasoning in the Self-organizing Information-Base System," *Computer Aided Innovation of New Materials* II, Elsevier Science Publishers B.V. (M. Doyama ed.), 1993, p.27-32.
13) Y. Fujiwara: "The Model for Self-structured Semantic Relationships of Information and Its Advanced Utilization," *International Forum on Information and Documentation*, Vol.19, No.2, 1994, p.8-10.
14) Y. Fujiwara and H. Gotoda: "Representation Model for Relativity of Concepts," International Forum on Information and Documentation, Vol.20, No.1, 1995, p.22-30.
15) H. Ishizuka, Y. Inoue, K. Oda, K. Katoh, Y. Kawashima, K. Shimizu, Y. Naka, M. Shiroki, T. Sekiya, H. Takashima, T. Nakayama, T. Nishioka, N. Hara, S. Sugawara, I. Matsuura, T. Miyagishima, M. Miyamura, R. Kani, S. Yoshida, N. Ohbo, H. Kitagawa, K. Yamakuchi and Y. Fujiwara: "Development of Computer/CD-ROM Aided Organic Synthesis Research System: CORES (In Japanese)," *13th Symposium on Chemical Information and Computer Science*, 1990, p.13-16.
16) J. Gasteiger and C. Jochum: "An algorithm for the perception of synthetically important rings," *J. Chem. Inf. Comput. Sci.*, Vol.19, 1979, p.43-48.
17) Edward H. Sussenguth, Jr: "A graph-theoretic algorithm for matching chemical structures," *J. Chem. Doc.* , Vol.5, 1965, p.36-43.
18) Peter Willett: "Similarity and Clustering in Chemical Information System," Research Studies Press Ltd., England, 1987.
19) M.A. Johnson and G.M. Maggiora: "Concepts and Applications of Molecular Similarity," Wiley, New York, 1990.
20) John M. Barnard: "Structure Searching Methods: Old and New," *J. Chem. Inf. Comput. Sci.*, Vol.33, No.4, 1993, p.532-538.
21) C. Berge: "Hypergraphs," North-Holland, 1989.

(Received February 8, 1996)

About the authors

Jianghong AN (Student Member)
He is a doctoral candidate in computer science at the University of Tsukuba.
E-mail: ajh@dblab.is.tsukuba.ac.jp

Yuzuru FUJIWARA (Member)
Dr. Fujiwara is a professor in computer science at the University of Tsukuba. Readers can contact him at the Database Laboratory, Institute of Information Science and Electronics, University of Tsukuba.
E-mail: fujiwara@dblab.is.tsukuba.ac.jp