

類推機能を持つ有機合成研究用情報ベースシステム

王忠清 鄭四清 藤原譲

筑波大学 電子・情報工学系

類推機能を持つ自己組織型情報ベースシステム、OS-IBS(Organic Synthesis Information-Base System)について報告する。類推は概念の類似性に基づく推論であり、概念間の類似関係は概念構造によって表わすことにし、推論を行なうための概念間の論理関係は論理構造によって表現することにした。このような情報の構造化はデータベースの内部情報を組織化することより実現された。類推の機能は構造化された情報空間に基づいて、概念間の類似性と論理関係によって実現された。有機合成研究用の情報を収録管理するデータベースシステムCORESがOS-IBSの情報源として利用された。

Analogical Reasoning in the Information-base System for Organic Synthesis Research

Zhong Qing Wang, Si Qing Zheng, Yuzuru Fujiwara

Institute of Information Sciences and Electronics, University of Tsukuba,
Tennohdai, Tsukuba, Ibaraki 305, Japan

This paper reports the self-organizing information-base system called OS-IBS(Organic Synthesis Information-Base System) with analogical reasoning function. The analogical reasoning is the reasoning based on the analogies between concepts. The analogical relationship between concepts is represented by the conceptual structure and the logical relationship between concepts used to realize the reasoning is represented by the logical structure. Such structuralizations of information are realized by self-organizing of the information existed in the database. Analogical reasoning is realized by evaluating analogies between concepts in the structuralized information space. The source information of OS-IBS comes from a database system called CORES which has been developed to store and to manage information necessary for research of organic syntheses.

1. はじめに

有機合成研究用の情報は大量かつ複雑である。膨大な量の化合物の間の有機反応は大きな情報空間を構成している。多量の有機合成に関するデータを収集、整理しデータベース化すること、およびそれを利用して新しい反応を予測し、また合成最適経路を設計するなど高度な応用システムを開発することは早くから期待されていた。化学情報に関するデータベースや知識ベースの構築が以前からの研究課題となっている。その中で、Corey「1」「2」をはじめとして、多くのシステムが開発されていた。これらを大別すると、反応規則を記述するエキスパート・システム型と反応事例を記述するデータベース型の二つになる。しかし、前者は反応ルールの収集、体系化が困難であり、後者は収録されていない反応に対処できないという問題点がある。

データベースの大規模化の問題点は、管理のため収録データに厳しい制約が課せられることである。その結果、研究開発用とか戦略決定用などの高度な情報を適切に資源化することは極めて困難である。また、現在ほとんどのデータベースと知識ベースシステムでは情報の利用機能として検索や演繹推論しか提供していない。情報を実際に利用する場合には、求められるデータが必ずしも収録されていないため、検索だけでなく既存のデータおよび知識から新しい知識を獲得する機能、つまり学習、類推などの機能が求められている。化学研究者のそのような要求を満足させるためには新しい機能の整備が必要となる「11」「12」。

そこで、IBS:SORITES(Information-Base Systems with Self-Organizing Receptor Interconnections)と呼ばれる新しい情報モデルが提案された「13」。このモデルにおいては、リンクが意味を持つことができるため、リンク情報を利用して、情報空間を構造化し機械学習や類推などを実現することができる。

類推というのは概念間の類似性に基づく推論である。類推機能は問題解決や予測推定において有力な手段であり、また計算機による情報、知識の処理をより柔軟に行なうためにも欠くことのできない機能である。類推に関しては、Winstonを代表として数多くの研究者により研究が行なわれている「3」「4」「5」「6」「7」「8」。しかし、多くの研究では、情報がすでに構造化されたという前提で研究を展開して行っており、情報をどのように構造化するかについてはあまりふれていない。これに対して、IBS-SORITESモデルでは、情報の構造化の程度が類推に決定的な影響を及ぼすという認識を持って、まず、データベースの所載の情報を、自己組織的に情報所在の位置関係を表わす物理構造、概念間の類似関係を表わす概念構造、と推論を行なうための論理関係を表わす論理構造に構造化させ、そして、類推を行なう。本論文では、IBS:SORITESのモデルに基づいて有機化合物の合成研究支援システムとして類推などの機能を持つ有機合成情報ベースシステム(Organic Synthesis Information-Base System. 以下OS-IBSと省略する)の開発、特に類推機能の実現の手順について報告する。

2. 有機合成情報ベースシステムの構成

有機合成情報ベースシステムは、有機合成研究に必要な情報を収録管理するデータベースシステムCORESE「15」「16」を情報源として利用している。情報ベースシステムに於ける類推等の機能はCORESにあるデータと知識に基づいて実現されたものである。

CORES(Computer/CD-ROM Aided Organic Synthesis Research System)は有機合成に関するデータを収集、整理してデータベース化すること、およびそれを利用して反応ルール等の知識を抽出し、知識ベースを作成したり、さらに有機合成最適経路設計などの高度な応用システムを開発することを目的として、分子設計研究会において作成されている。データセットとその内容、表現形式およびシステムの基本機能と拡張性から、色々な形でルールや知識の抽出、種々の有機合成反応研究支援システムへの応用が可能である。CORESデータベースには有機合成の総説である"Organic Synthesis"の第1巻から第68巻などの全文データが収録されている。そして、

1) "Organic Synthesis"にある索引データ。索引データは一般索引、反応タイプ索引、化合物索引、CAS名索引、分子名索引、著者索引などを含んでいる。これらの索引には専門家が重要と思う有機合成の概念が含まれている「14」。

2) 専門家の作成した化学反応の要約データ。要約データには化学反応における出発物や生成物、化学反応を行なうための試薬、温度、溶媒、反応時間および生成物の収率に関する情報を記述している。要約データは、FROM reactant WITH reagent IN solvent AT temperature Y=yieldのフォーマットを持っている。

TITLE: BENZOYLENE UREA

LITERATURE: ORGANIC SYNTHESSES COL. VOL. 2 P. 79

SUMMARY: FROM ANTHRANILIC ACID AND KCNO WITH ACOH IN H₂O AT 30. Y = 85%

図1 要約データの例

3) 化合物の構造データ。

4) 構造検索用検索コードデータ

などの情報も入っている。

OS-IBSでは、CORESにある"Organic Synthesis"の索引データから得られた有機合成に関する重要概念を含む索引情報と化学反応要約データから得た官能基、試薬、反応条件等の知識を持つ関係データベースを用いて、化合物間の論理関係を表すリンクを自己組織的に生成し、情報の構造化を実現した。類推等の機能がこの構造化された情報空間の上で実現された「9」「10」。

全文情報のような一次情報、書誌情報のような二次情報、図、表のような多次元情報などを含んでいる情報ベースシステムの構成は図2に示されている。

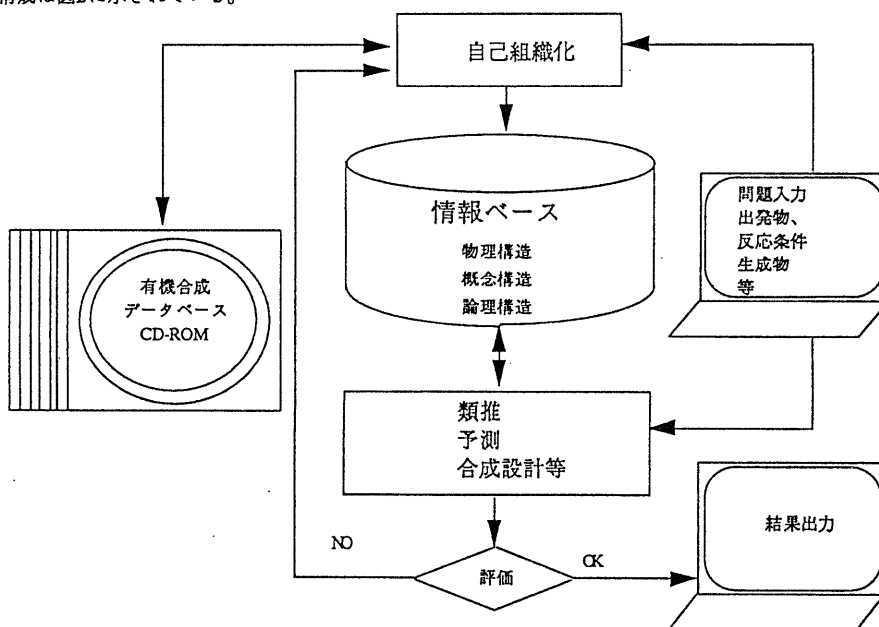


図2 情報ベースシステムの構成

3・情報の構造化

3・1 情報の物理構造

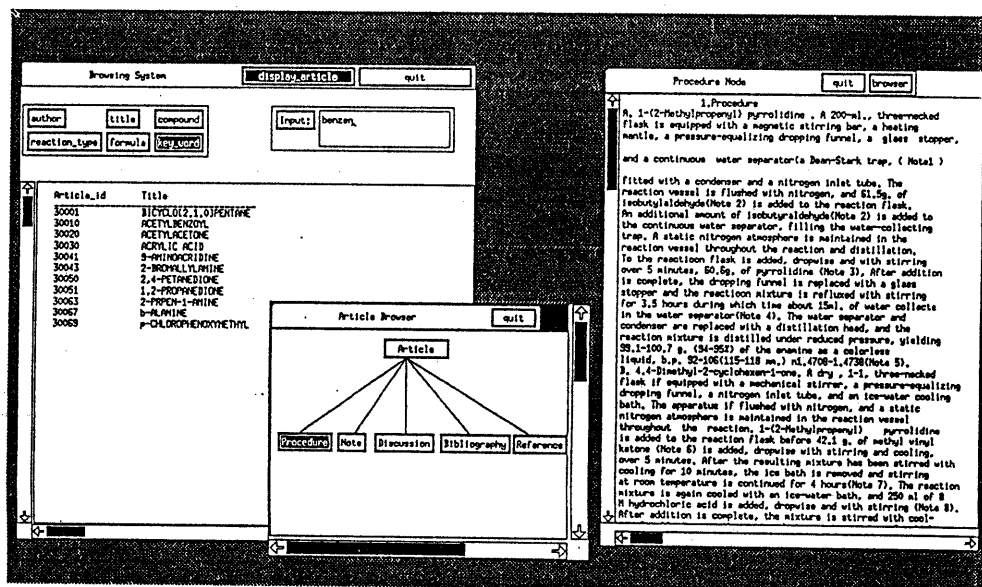


図3 物理構造

情報をアクセスするための、情報を記録する位置を表わすのは情報の物理構造である。CORESに含まれている"Organic Synthesis"は個々の文献から構成されている。各文献では有機化合物の生成方法、生成過程及び生成時に必要な器具等の情報が記述されている。また各文献は、(1)文献のタイトル、著者などに関する書誌情報、(2)化学反応式、(3)反応過程(Procedure)、(4)反応過程で使われている化合物に対する注釈(Notes)、(5)文献全体に対する考察や他の研究との比較(Discussion)と(6)参考文献(References)という六つの部分から構成される。印刷物情報に対応する物理的構造は図3に示している。

3・2 情報の概念構造

化合物の概念構造

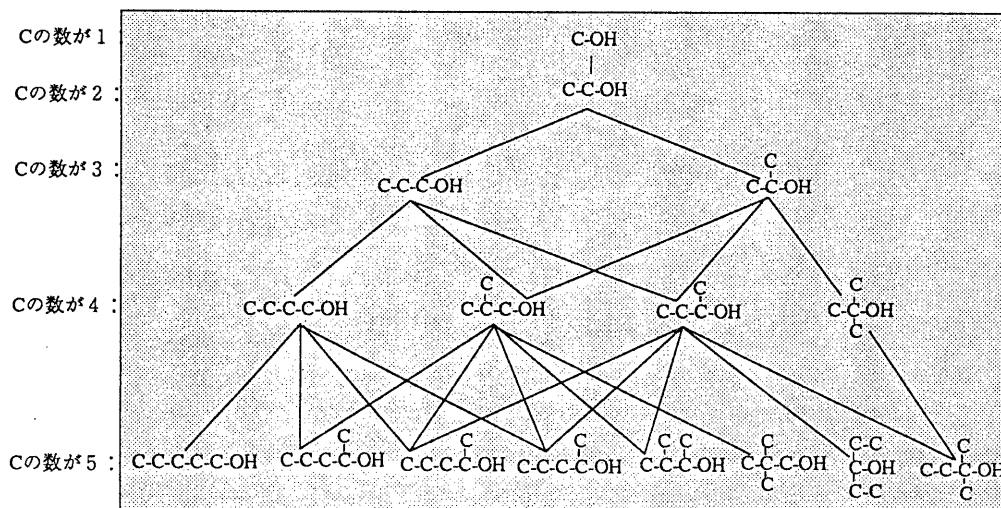


図4 概念構造の一例

概念構造とは概念間の類似関係を表すものである。図4では炭素の増減に対応するアルコール概念の間の類似関係を表す概念構造の例が示されている。ここでは、線は炭素が一つ追加、削減された化合物の関係を表している。概念間の類似性はこのような概念間の遠近関係によって示される。概念構造を利用して、ある概念と近い概念を見出すことができる。情報ベースシステムの類推機能において、類似概念の探索はこの概念構造に基づいて行なわれる。

3・3 情報の論理構造

QS-IBSにおける類推は概念間の類似性に基づく推論である。推論を実現させるための概念間の論理関係を表わすのは論理構造である。CORESデータベースの索引情報とサマリーデータを用いて、反応規則、すなわち化合物間の論理関係を記述するリンク情報を自己組織的に抽出するサブシステムを開発した「9」「10」。リンク情報としては、Link_id、type(反応タイプ)、label(反応パターン)、reactant、reactant_formula、reactant_location、reagent、temperature、solvent、time、yield、product、product_formula、product_location等がある。このリンク情報を利用して、有機合成情報の論理構造を構成することができる。

図5は還元の論理構造、 $\text{RCH}_2\text{OH} \rightarrow \text{RCHO}$ という反応パターンの論理構造と $\text{RCH}=\text{CH}_2 \rightarrow \text{RCH}_2\text{CH}_3$ という反応パターンの論理構造の例を示している。

酸化反応タイプの論理構造		
出発物	生成物	Link_ID
CH ₃ CH ₂ CH ₂ CH ₂ OH	-> CH ₃ CH ₂ CH ₂ CHO	----- (1)
CH ₃ CH ₂ CH ₂ OH	-> CH ₃ CH ₂ CHO	----- (56)
CH ₃ CH ₂ CH ₂ CH=CH ₂	-> CH ₃ CH ₂ CH ₂ CH ₂ COH	----- (32)
CH ₃ CH ₂ CH=CH ₂	-> CH ₃ CH ₂ CH ₂ COH	----- (20)
CH ₃ CH=CH ₂	-> CH ₃ CH ₂ COH	----- (6)

RCH ₂ OH->RCHOの反応パターンの論理構造		
出発物	生成物	Link_ID
CH ₃ CH ₂ CH ₂ CH ₂ OH	-> CH ₃ CH ₂ CH ₂ CHO	---- (1)
CH ₃ CH ₂ CH ₂ OH	-> CH ₃ CH ₂ CHO	--- (56)

RCH=CH ₂ ->RCH ₂ COHの反応パターンの論理構造		
出発物	生成物	Link_ID
CH ₃ CH ₂ CH ₂ CH=CH ₂	-> CH ₃ CH ₂ CH ₂ CH ₂ COH	---- (32)
CH ₃ CH ₂ CH=CH ₂	-> CH ₃ CH ₂ CH ₂ COH	--- (20)
CH ₃ CH=CH ₂	-> CH ₃ CH ₂ COH	--- (6)

図5 論理構造の例

conditions for reaction search

reactant

CH₃CH₂CH₂CH₂OH

product

CH₃CH₂CHO

reaction_type

oxidation

reaction_pattern

ROH->RCHO

search

analogical reasoning

There are no such reactions.

The reactions with the same reaction_pattern have been searched out:

REACTANT	PRODUCT	LINK_ID
CH ₃ CH ₂ CH ₂ CH ₂ OH	-> CH ₃ CH ₂ CH ₂ CHO	---- (1)
CH ₃ CH ₂ CH ₂ OH	-> CH ₃ CH ₂ CHO	----- (56)

The input reaction may be carried out under the conditions similar to those of the above two reactions because the following reactions are carried out under almost the same conditions.

REACTANT	PRODUCT	LINK_ID
CH ₃ CH ₂ CH ₂ CH=CH ₂	-> CH ₃ CH ₂ CH ₂ CH ₂ COH	----- (32)
CH ₃ CH ₂ CH=CH ₂	-> CH ₃ CH ₂ CH ₂ COH	----- (20)
CH ₃ CH=CH ₂	-> CH ₃ CH ₂ COH	----- (6)

図6 検索の例

4・ 類推による反応ルールの制御

OS-IBSでは、構造化された情報空間の上で、類推機能、すなわち、概念構造上の概念の類似性に基づく推論を利用して、ルールの制御を行なっている。有機合成において専門知識の程度は反応規則の適用についてその可否を精度よく予測できることで示されている。すなわち反応パターンを知ることよりその適用可能な範囲と限界を正確に把握することがより重要であり、また困難である。「9」では、類推による反応ルールの制御の可能性について説明したが、具体的な手順については言及しなかった。そこで、以下では、概念構造と論理構造に基づく類推によるルール制御の具体的な手順を説明し、5では実例を示す。

有機合成情報ベースシステムでは図6で示すように出発物の分子式、生成物の分子式、反応タイプと反応パターンから合成反応を検索することができる。検索の結果はウィンドウ上に表示される。検索しようとするものがデータベースになれば、システムの類推機能を用いて、この反応と同じような条件で進むことのできる反応が類推され、ウィンドウ上に表示される。類推の具体的な手順は以下で説明する。

4. 1 反応条件の類似性の判定基準

データベースから検索された反応に対して、反応の条件が同じかどうかの判定はリンク情報、すなわち、試薬、温度、時間および収率等に基づいて行なわれる。具体的に言うと、二つの反応において、もし二つの反応の試薬が同じで、反応の温度の差が50℃以内で、収率の比は50%と200%以内であれば、この二つの反応条件が同じであると判定する。

4. 2 類似反応の検索

OS-IBSにおける類推機能は、まず3. 2で説明した概念構造を利用して、未知の反応の性質と最も近い反応を探し出し、そして、それらの反応を評価することによって、行なわれる。概念構造の中で、距離が近いほど化合物の性質が似ていて、当然反応条件も距離の近いものより近いという前提に基づいて、化合物の概念構造を利用し、検索しようとする反応と類似する反応を検索する。具体的には、検索しようとする反応と同じ反応パターンでかつ出発物と生成物の炭素(C)の数が検索しようとする反応の出発物と生成物の炭素の数より一つ少ないあるいは多い反応を検索する。もしなければ、より二つ少ないあるいは多い反応を検索する。……このようにして、検索しようとする反応と同じパターンである反応を見つけるのである。

4. 3 類似性の判断

類似した反応の反応条件が検索しようとする反応に当て嵌めることができるかどうかを判断する必要がある。判断の方法は以下である。

1) まず、反応パターンを用いて、類似性を判断する。4. 2で検索された反応が複数である場合、これらの反応の条件を類似性の判定基準に基づいて比較する。反応条件が同じであると言えれば、これらの反応と似ている条件で、検索しようとする反応を実行することができるとの判断を下す。

2) 反応タイプを利用して、類似性の判断。4. 2で検索された反応が一つしかない場合、同じ反応タイプで、違った反応パターンを持つ反応を検索し、類似性を判断する。すなわち、次のような条件を満たす反応を検索する。(1) 検索しようとする反応と同じタイプである。(2) 出発物と生成物の概念構造は検索しようとする反応の出発物と生成物の概念構造と同じある、あるいは4. 2で検索された反応の出発物と生成物の概念構造と同じである。具体的な判断方法は以下である。まず、反応タイプが検索しようとする反応と同じでかつ出発物と生成物の変化していない部分(官能基を除いた部分)が検索しようとする反応の出発物と生成物の変化していない部分と同じである反応を検索する。次に、同じ反応タイプでかつ出発物と生成物の変化していない部分(官能基を除いた部分)が4. 2で検索された反応の出発物と生成物の変化していない部分と同じである反応を検索する。このように検索された、検索しようとする反応と違った反応パターンを持つ反応の条件を反応類似性の判定基準に基づいて調べる。もしこれらの反応条件が同じであると言えれば、4. 2で検索された反応と似ている条件で、検索しようとする反応を実行することができるとの判断を下す。

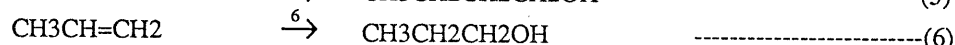
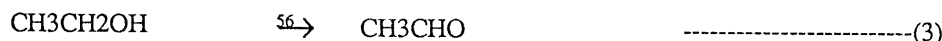
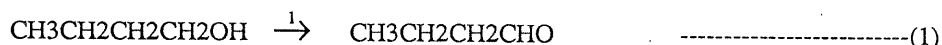
反応の類似性を1)で判断された後、類推の精度を高める目的で、2)でさらに評価することができる。5. 1ではこの例を示す。

4. 4 類推結果の表示

4. 2で検索された類似反応は、4. 3で示された方法によって評価され、類推の結果はウィンドウ上に表示する。

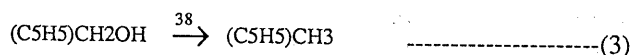
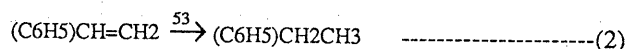
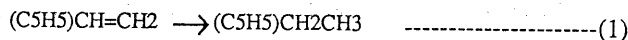
5・類推の実例

5・1 反応ルールが応用できる類推



データベースに存在していない反応を検索する場合、類推機能を利用して、この反応の反応条件を予測することができる。この例では、類推を使ってデータベースに存在しない反応2の反応条件を予測することを示す。図4で示された概念構造を利用して、CH₃CH₂-と距離が一番近いものはCH₃-, CH₃CH₂CH₂-と(CH₃)₂CH-であることがわかる。すなわち、炭素の数がCH₃CH₂より一つ少ないあるいは多いものの方が一番近い。したがって、情報ベースシステムでは、まず、反応2と同じ反応パターンで(RCH₂OH>RCHO)でかつ炭素の数が反応2より一つ少ないあるいは一つ多い、反応1と反応3を探し出した。反応1と反応3の反応条件はそれぞれ1番目のリンクの56番目のリンクに入っている。この二つのリンクに入っている反応情報は4. 1で説明した類似性判定基準に満たすため、反応1と反応3の条件を反応2に当て嵌めることができるか否かを判断するため、反応1と反応3に似ている条件で反応2を実行することができるのと類推結果を出すことができる。この例では、類推の結果を更に高める目的で、4. 2の2)番目の判断方法を利用して、反応4、反応5と反応6を探し出した。反応4、反応5と反応6は反応パターンが同じで、反応タイプが反応1と同じで(酸化反応)である。反応4の変化していない部分は反応1と同じくCH₃CH₂CH₂で、反応5の変化していない部分は反応2と同じくCH₃CH₂で、反応6の変化していない部分は反応3と同じくCH₃である。反応4、反応5と反応6の条件がそれぞれ32番目のリンク、20番目のリンクと6番目のリンクに入っているため、32番目のリンク情報、20番目のリンク情報と6番目のリンク情報を4. 1の判定基準に基づいて評価した。この結果としては、この三つの反応条件が似ていることであった。すなわち、4. 2の2)番目の判断方法を利用しても、同じく、反応1は反応2と反応3に類似する条件で行なうことができるのではないかと類推の結果が出された。

5・2 反応ルールが応用できない類推



例1では、類推を使ってデータベースに存在しない反応の反応条件を予測することを示したが、例2では、反応ルールを不正に適用するのを除外することを示す。反応条件が予測されようとする反応1に対して、概念構造上で距離が近

い反応は反応2しか検索できなかったため、更に例1と同じ手順で反応3と反応4が探し出された。しかし、反応3と反応4の条件が入っている38番目のリンク情報と36番目のリンク情報を4・1の判定基準に基づいて評価した結果としては、この二つの反応条件が異なったため、反応2の条件は反応1に適用できないという類推の結果が出された。この例では、C6H5-とC5H5-が概念構造上で一番近いに関わらず、性質が違うため、反応ルールを不正に適用しては行けないことを示した。

6・結論

類推機能を持つ自己組織型有機合成研究用有機合成情報ベースシステムについて報告した。類推は概念間の類似性に基づく推論である。概念間の類似関係は概念構造で表われ、推論を行なうための概念間の論理関係は論理構造で表われる。このような情報の構造化は自己組織的に行なうことができる。類推機能による反応ルールの制御の具体的な手順および実例を示した。

参考文献

- 1) E. J. Corey, W. T. Wipke: Computer-assisted Design of Complex Organic Syntheses. Science, Vol. 166, pp.178-192(1969).
- 2) E. J. Corey, A. K. Long, S. D. Rubinstein: Computer-assisted Analysis in Organic Synthesis. Science, Vol. 228, pp.408-418(1985).
- 3) R. P. Hall: Computational Approaches to Analogical Reasoning, A Comparative Analysis. Artificial Intelligence, Vol. 39, pp. 39-120(1989).
- 4) S. Kedar-Cabell: Analogy - From a Unified Perspective, in: D. H. Helman (Ed.). Analogical Reasoning: Perspectives of Artificial Intelligence, Cognitive Science and Philosophy, pp. 65-103 (1988).
- 5) Makoto Haraguchi and Setsuo Arikawa: A Formulation of Analogical Reasoning and Its Realization(In Japanese), Journal of Artificial Intelligence Society, Vol.1, No.1, pp. 132-139 (Sept. 1986).
- 6) Masayuki Numao and Masamichi Shimura: Analogical Reasoning by Decomposing Explanation Structure(In Japanese), Journal of Artificial Intelligence Society, Vol.6, No.5, pp. 716-724 (Sept. 1991).
- 7) Patrick H. Winston: Learning and Reasoning by Analogy, Communications of the ACM, Vol. 23, No.12, pp. 689-703(1980).
- 8) Patrick H. Winston: Learning New Principles from Precedents and Exercises, Artificial Intelligence, Vol. 19, pp. 321-350 (1982).
- 9) Zhong Qing Wang, Si Qing Zheng, Xu Yu, Kazunori Yamaguchi, Hiroyuki Kitagawa, Nobuo Ohbo and Yuzuru Fujiwara: Learning and Analogical Reasoning in The Information-Base System for Organic Synthesis Research, Journal of Japan Society of Information and Knowledge, Vol.2 No.1, pp. 93-99(1991).
- 10) Zhong Qing Wang, Naohiro Hayakawa, Kazunori Yamaguchi, Hiroyuki Kitagawa, Nobuo Ohbo and Yuzuru Fujiwara: Learning and Rule Control in the Information Base System for Organic Synthesis Research, 14th Symposium on Chemical Information and Computer Science, pp. 162-165, Nov., 27-29, (1991).
- 11) Yuzuru Fujiwara, Jihong He, Gyoto Chang, Nobuo Ohbo, Hiroyuki Kitagawa and Kazunori Yamaguchi: Self Organizing Information-Base Systems for Material Design, Proc. of the 1st CAMSE, Tokyo, (Aug., 1990).
- 12) Yuzuru Fujiwara, Nobuo Ohbo, Hiroyuki Kitagawa and Kazunori Yamaguchi: The Information-Base Systems for Materials Research, The 12th CODATA Conference, Columbus, Ohio, (July 15-19, 1990).
- 13) Y. Fujiwara: The Self organizing Information-base Systems with Learning and Analogical Reasoning, Proc. of the 3rd Beijing Int. Symp. on Computer. Information Management, S3A-7, (Oct. 14-Oct. 18, 1991).
- 14) W. E. Noland (ed): Organic Syntheses, Collective Volume 6, John Wiley & Sons (1988).
- 15) Ikutoshi Matsuura: Reaction Data Base, Molecular Design, Vol. 7, No. 5, pp. 2-33 (1986).
- 16) 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. Yamaguchi and Y. Fujiwara: Development of Computer/CD-ROM Aided Organic Synthesis Research System: CORES(In Japanese), 13th Symposium on Chemical Information and Computer Science, pp. 13-16, (1990).