

有機合成反応のDB/KB統合化システム

—動的知識の自動抽出について—

伊藤照明、大保信夫、藤原 譲 (筑波大学)

中山堯 (国際科学振興財団)

3J-8

有機合成反応データベース及び知識ベースの統合化による反応設計支援システムの構築を行っている。(1) 本稿では、システムの中核をなす動的知識自動抽出のためのサブシステムについて報告する。

【DB/KB統合化と知識の自動抽出】

一般に、こうしたシステムを構築する上でまず第一に考えるべきことはデータベースの充実である。本システムでは、個別反応の出典として有機合成研究者のバイブルともいえる Organic Synthesis を選んだ。ここに収録された数多くのデータは全て追跡実験により評価されている信頼性の高いデータである。グラフィックエディター(2)により出発物の構造式を入力し、その部分構造を修正することで反応生成物の構造とする。そのようにして原子の対応をつけた出発物と生成物とを結合表と座標のデータに変換したものを Reaction Table とし、それに反応の要約を付け加えることで個別の反応を表現する。Organic Synthesis 全65

巻に収録されている反応データをこうした方法で表現し、蓄積してCD-ROM化したものが本サブシステムの対象となるデータベースの主要部を構成している。

ところで、現在のエキスパートシステムの抱えている大きな問題点の一つはその知識獲得にある。本システムではデータベースからの知識の自動抽出により知識獲得を行う。つまりデータベースに格納された個別の反応事例から反応ルールの自動抽出により知識ベースが構築される。

システムにおける反応の内部表現と、それから知識を獲得する方法の例を図1に示す。ここで、o-トルアミドの合成反応に関する情報はグラフィックエディターを介して Reaction Table としてデータベースに格納されている。この Reaction Table は出発原料と生成物の結合表を簡易結合表に変換して合成して得られたものであり、その逆変換をすることで本来の結合表を得ることができる。そして、これらの結合表の差をとることで部分構造変化つまり反応のルールの抽出を行うことができる。

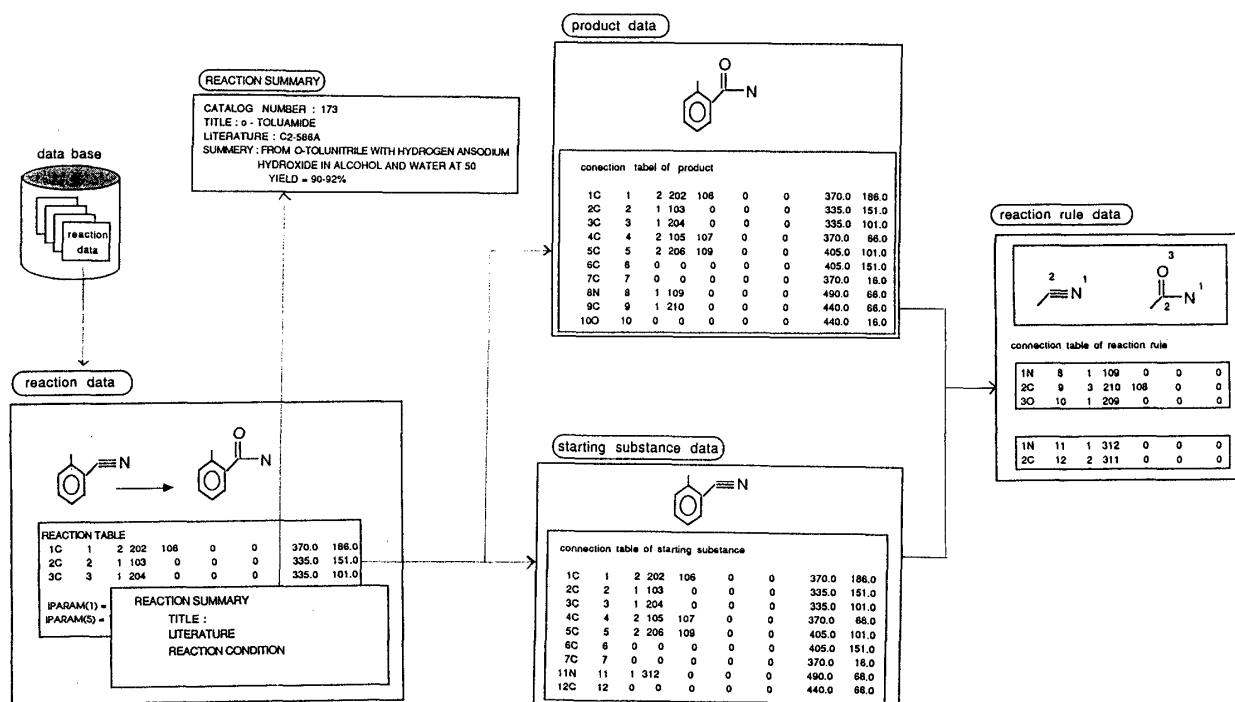


図1 o-TOLUAMIDE 合成における反応ルールの抽出方法

DB/KB Integrated System for Organic Synthesis

Teruaki ITOH\*, Takashi NAKAYAMA\*\*, Nobuo OHBO\*\*\*, Yuzuru FUJIWARA\*\*\*

\* Division of Scientific Technology, Univ. of Tsukuba

\*\* International Foundation for Advancement of Science

\*\*\*Inst. of Information Sciences and Electronics, Univ. of Tsukuba

この例の場合には Reaction Table 及び付加情報の解析を行うことで反応に関して次のような知識を得ることができる。

- 特徴量 出発物 1 : ベンゼン環 ニトリル基(-CN)
- 生成物 2 : ベンゼン環 アミド基(-CONH<sub>2</sub>)
- 反応ルール : ニトリル基(-CN)からアミド基(-CONH<sub>2</sub>)  
                  への変化
- 反応条件 : 触媒 H<sub>2</sub>O<sub>2</sub>+NaOH  
              溶媒 alcohol+H<sub>2</sub>O  
              反応温度 50°C
- 反応収率 : Y = 90-92%
- 出典 : Organic Synthesis col. vol.2 p586

このような情報がデータベースから自動的に抽出される。

#### 【動的知識の抽出】

このようにして得られた反応ルールをそのまま化合物に適用しても、推定された反応生成物または前駆体は、立体障害や反応条件等様々な理由で実際の化学反応では得られないような物質である可能性がある。つまり、このような知識を適用して得られた結果は必ずしも化学研究に有用な情報とは限らないのである。これは適用した知識が汎用的ではあるが精度が低いためと考えられる。こうした汎用的な知識はもちろん必要であるが、化学研究者の要求を満足できるシステムとするためにはできるだけ精度の高い知識を、つまりできるだけ個別の反応に近いルールを適用することが有効であると思われる。ところが、部分構造の変化部のみに着目した上述の知識ではそのような対応をすることができない。そこでそうした対応を可能とするために、“動的な知識”という知識の新しい表現方式を導入した。個別の反応について、specific なレベルから generic なレベルまでのいろいろなレベルで知識獲得を行い、それらを合成して化学反応を表現するものである。“動的な知識”を用いたシステムは、specific なレベルで最も精度の高い知識を適用することをまず行い、それが適用できない場合には、少しずつ generic なレベルにした知識で対応

していくわけである。そうしたレベルでも適用できる知識がない場合には、最も generic なレベルの知識で対応していくことになる。図1 で得られた知識はこの最も generic なレベルの知識に相当する。本研究ではReaction Table からこのような動的知識の自動抽出を行えるサブシステムを開発した。動的な知識の抽出例を図2 に示す。

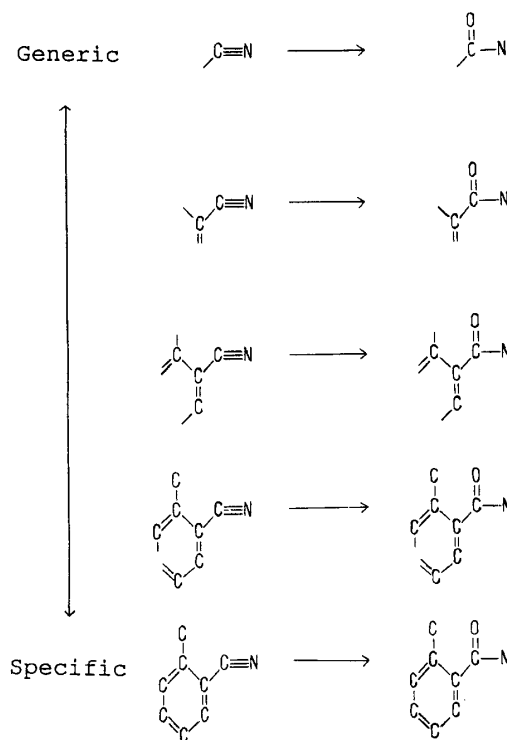


図2 動的な知識抽出の例

#### 【結果と考察】

汎用のエキスパートシステム構築用ツールは主として符号のデータやルールを扱うので、本システムのように化合物という対象を扱うシステム構築には不十分な点が多い。そこで、本研究では化学反応に適したツール開発をも含めたシステム開発を行っている。本研究では有機合成反応のデータベースから動的知識の自動抽出を目的としたサブシステムの開発を行い有意義な結果が得られた。この“動的な知識”を生かした推論システムの開発についても述べる。

#### 【参考文献】

- 1) 伊藤、藤原他“化合物のDB/KB統合システム”、第36回情報処理学会全国大会3P-5.
- 2) 中山、藤原他“合成経路選定における動的仮設定について”、第9回情報化学討論会論文集(1986, 10)
- 3) 中山、藤原他“化学反応設計支援システム用知識ベースの設計”、第33回情報処理学会全国大会5M-4.