

電子投稿による
SGML方式全文データベースの作成
— 日本化学会の実験 —

○石塚英弘* (図書館情報大)、伊藤 卓 (横国大工)、竹内敬人 (東大教養)、
千原秀昭 (化情協)、中野英彦 (姫路工大工)、眞野倅一 (信州大教養)、
吉村忠与志 (福井高専)、中西敦男 (日本化学会)、田中洋一 (凸版印刷)

* E-mail: ishizuka@ulis.ac.jp

93年1月号からSGML方式全文DBによる電子出版になった日本化学会欧文誌の電子投稿の研究である。94年4月、日本化学会は広く使用されているワープロソフトMS Word のスタイルシート機能を使った電子投稿の実験を開始した。この手法は、1)投稿者がSGMLを意識させない、2)交換フォーマットであるRTF(Rich Text Format)を使用するため、プログラムでSGML方式全文DBに変換できる、3)他のワープロソフト、WordPerfect, PageMakerでも投稿可能、などの長所を持つ。

Generation of Full-text Database
Based on SGML through
an Electronic Contribution by an Author
An experiment in Chemical Society of Japan

Hidehiro Ishizuka (Univ. of Library and Information Science);
Takashi Ito (Yokohama National Univ.); Yoshito Takeuchi (Univ. of Tokyo);
Hideaki Chihara (Japan Assoc. Internat. Chemical Inf.);
Hidehiko Nakano (Himeji Inst. Tech.); Koichi Mano (Shinshu Univ.);
Tadayoshi Yoshimura (Fukui Polytech.); Atsuo Nakanishi (Chem. Soc. Japan);
Yoichi Tanaka (Toppan Printing Co. Ltd.).

This paper reports a method of electronic contribution by an author to "Bulletin of the Chemical Society of Japan", that has been published with a SGML-based system since January 1993. Chemical Society of Japan started an experiment of electronic contribution using a style sheet function and RTF (Rich Text Format: an exchange format) of Microsoft Word, which is one of the most popular word-processing software. This method has the following features; (1) an author need not know SGML; (2) contributed file in RTF can be converted to SGML-based full-text database by a program; (3) the method can be used in other word processing software: WordPerfect or PageMaker.

1. はじめに

最近日本でもSGML方式による全文データベース(DB)の重要性が認識されるようになってきた。しかし、実際にDBを作成するには未だ解決すべき課題が残されている。特に、異なる組織に属する人々の間でDBを構築する場合は、検討が必要である。本稿では、その例として学会への論文投稿を取り上げ、ワープロソフトのスタイルシート機能を使った電子投稿とSGML方式全文DBの生成の実験を紹介し、その手法と有効性を示す。

1.1. SGML方式の浸透

欧米ではSGML(Standard Generalized Markup Language)^{1) 2)}が全文DB作成・管理手法として様々な分野で広く使われている。SGML方式は、1)テキストだけでなく、図表、数式、写真等々、印刷される全てのデータを扱えること、2)章、節、パラグラフなどの階層構造が表現できること、3)図表参照や文献参照など参照関係(非階層構造)も表現できること、などの特長を持つ。そのため、特にマニュアルや製品仕様書など技術文書の作成・保守・情報処理でよく使われている。中でも、1988年に始まったCALSは有名である。CALSは当初、米国国防省の文書電子化プロジェクト、Computer aided Acquisition and Logistics Supportの略としてスタートしたが、93年からはContinuous Acquisition and Life-cycle Supportの略に変更され、民生用、即ち、一般官庁や民間企業間の取引にも適用されるようになってきた。そして、CALSはNII(情報ハイウェイ)の上に乗り、EC(電子商取引)とも結びつこうとしている。こうなると、製品仕様書、各種技術文書、設計、生産、発注、決裁などがSGMLと結びついてくる。一方、学術雑誌への適用例を挙げれば、米国マサチューセッツ医学会によるAIDSの教科書"The AIDS Knowledge Base"(AIDS-KB)、89年にElsevierがSGMLの採用を決めたことなどがある。

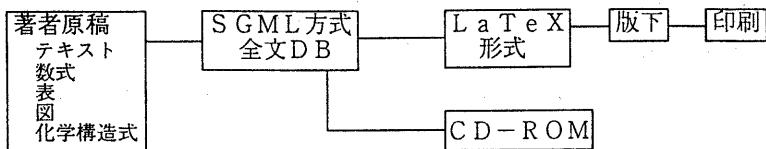
日本でも最近SGMLが浸透してきた。たとえば、コンピュータや通信の業界では富士通がマニュアル等技術文書へのSGML導入を進めており、他の会社も続いている。また、VLSIのDBのSGML方式による構築が半官半民の組織で始まった。特許庁のシステムでも特許情報のDBがSGML方式になった。厚生省は95年度から新薬の登録書にSGML方式を導入することを決めた。学術雑誌の例³⁾を挙げれば、90年12月の情報知識学会誌³⁾のSGML方式電子出版に始まり、学術情報センターによる91年3月の「SGML実験誌」⁴⁾等々があり、これらを経て日本化学会⁵⁾は1993年1月号から欧文論文誌のSGML方式全文DBを作成し、印刷している。一方、CALSも注目されるようになり、94年9月の晴海のデータショウにCALS Japan '94が併設された。

SGMLが浸透してくれれば、SGML方式システムに提出する時のみSGML方式DBを作るよりも、それ以前の段階からSGML方式に対応できるようにした方が効率的である。製品の仕様書やマニュアル、特許など個別段階で作るのではなく、研究開発段階から対応した方がよい。論文やレポートのSGML対応も進める必要があろう。

1.2. 電子投稿

電子投稿には、印刷工程の合理化を狙うものと、SGML方式全文DBの構築への寄与を狙うものがある。後者では、印刷はDBからソフトを使って行う。前者の例としては、88年からのTetrahedron Computer Methodology (TCM誌、コンピュータ化学の論文誌)があり、最近では日本でも物理・数学系の学会でLaTeXでの電子投稿を採用した所がある。後者の例としては、前述のAIDS-KBや情報知識学会誌などがある。

SGML方式 (1993年1月号から)



電子投稿 (試行)

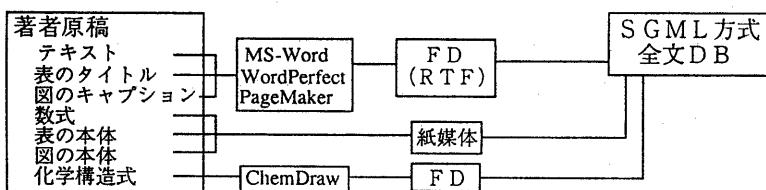


図1. Bull. Chem. Soc. Japanの電子出版工程

TCM誌は、米国 Molecular Design 社が開発した化学文書作成用ワープロソフト:ChemTextを使って編集・印刷を行い、投稿方式はフロッピーディスクによる電子投稿を基本とした。電子投稿可能なファイル形式は、ChemText, ASCII, MS Word (Macintosh)であった。

AIDS-KB、情報知識学会誌とともに、文書を構成する要素(たとえば、タイトル、章、章のタイトル、節、節のタイトル等々)を区別するための簡単な記号を設定し、その記号を目印としてソフトでSGML方式DBを生成する。この方式は、簡単なエディタ・ソフトでも文書を作ることができる点が長所である。しかし、文章中にタグが混在して若干見にくくなるという短所もある。

会社内で実施されるマニュアル作成やCALSでは、特定のソフトやフォーマットを指定することができるが、学術論文の場合は難しい。その理由は、1)投稿者は雑誌を選択できるため、ソフトやフォーマットを規制する雑誌は嫌われやすいこと、2)CALSや特許と異なり、利益と関係ない学術論文では投稿に高価なソフトを使うメリットはないこと、の2点が挙げられる。

そこで我々すなわち日本化学会の化学論文電子投稿検討委員会(伊藤卓委員長)は、スタイルシートを使う方法を採用した。それは、1)プログラムによりSGML方式全文DBに変換でき、2)広く使われているワープロソフト(MS Word, WordPerfect, PageMaker)で使用でき、3)文書作成中の見た目も良いからである。そして、94年4月からモニターを募ってBCSJの電子投稿の実験を開始⁶⁾した。

2. 日本化学会欧文誌のSGML方式全文DB

SGML方式全文DBは、その構造を記述した部分(DTD)とDB(文書)のデータそれ自体とから構成される。DTDはDocument Type Definitionの略で、DBの構造をSGMLの構文を使って定義・記述したものである。文書のデータには、DTDに従ってマーク付けされたテキスト(文書テキスト)と、図、化学構造式、反応式、写真などをビットパターン化したファイル(外部エンティティ)がある。外部エンティティもDTDに従っており、本文との参照関係は文書テキスト中に書き込まれている。なお、表、数式、分子式などは文字で書けるため、文書テキストの中に書くことができる。

以前の電算写植による印刷に変わって、日本化学会の欧文誌(BCSJ)は、1993年1月号から「SGML方式全文DBに基づく電子出版」になった⁵⁾(図1参照)。この方式では、まずSGML方式の全

文DBを作り、次にそれを基にして、印刷、情報検索用CD-ROMの作成、オンライン情報検索システム用DBへの変換など様々な電子出版を行う。図2に印刷例を示した。オンライン情報検索システムへのDB提供⁷⁾は、学術情報センター向けは実施済で、STN Internationalへは技術的検討は済み、現在は経営の観点から検討中である。また、94年10月末から12月末までWWW Mosaic版を試験公開している。

DBのデータ例を図3に示す。タイトル、著者名、抄録、セクションのタイトルなど各要素は、開始タグ(<要素名>)と終了タグ(</要素名>)で挟んで表現されている。また、上付はSUPで、下付はINFで、イタリックはITで表現する。なお、要素名とその関係はDTDで定義する。

なお、BCSJでは投稿者の混乱を避けるため、93年時点では投稿は従来どおり紙媒体による方法とした。そして、タグを示す簡単な記号を設定し、印刷会社の専門オペレータによって入力し、その記号をプログラムで正規のタグに変換して、全文DBを作っている。

3. スタイルシート機能・RTF形式とSGML方式全文DBへの変換

スタイルシート機能とは、センタリングの有無、書体、字の大きさ等々の書式機能を、タイトル、論文タイトル、セクション・タイトル、本文など各項目ごと個別に設定できる機能で、MS Word、WordPerfect、PageMakerなどのワープロソフトに用意されている。また、Microsoft社は、スタイルシート機能だけでなく、上付、下付も含む書式機能付きの文書ファイルを他のワープロソフトとの間で交換できるように、Rich Text Format (RTF)を設定した。RTFは他社にも受け入れられ、MS Wordだけでなく、WordPerfect、PageMakerなどのソフト間での交換が可能になった。

そこで演者らは、1)SGML方式の要素に対応するスタイルシートをRTF形式で作成し、そのフロッピーディスク(FD)を投稿者に配付する、2)投稿者はMS WordあるいはWordPerfectかPageMakerによって論文を作成し(図4参照)、RTF形式の原稿をFDで投稿する、3)RTF形式のファイル(図5参照)から変換プログラムによってSGML方式全文DB(図3参照)を作成する、の手順による電子投稿を計画した。図4に示した例はMacintosh上のMS Wordで作成したものであるが、Windows上のMS Wordでも同じ操作で作ることができる。表1にBCSJの電子投稿用スタイル一覧を示したので参考されたい。表中のContents with Illustrationsは化学構造式、反応式などの図を付けた目次で、94年から掲載することになったものである。なお、RTFで記述できない数式、表は従来どおり紙媒体で提出願うこととした。また、化学構造式、反応式はよく使われているChemDrawでの提出を期待することにした。現在、モニターを募り実験中である。

4. 本手法の評価・検討

この方法を採用した理由は、単に入力、校正の手間が減り、出版までの期間を短縮しうることだけでなく、1)項目が区別できるスタイルシート機能とRTFを使用したため、プログラムでSGML方式全文DBに変換できること、2)化学会でアンケート調査をしたところ、英文ワープロではMS WordやWordPerfectの利用者が多かったこと、3)これらワープロソフトの機能を使えば、上付、下付、イタリック、ボールドなどの指示も容易に伝わること、などがある。

94年10月初め現在、モニター数は約50人である。これまでのところ、前述した長所は実現し、評価されているが、彼らからは更に、1)ギリシャ文字は α の形式ではなく、 α と入力したい、

Bis(η^5 -methylcyclopentadienyl)dihydridomolybdenum(IV). Synthesis and Reactions with Nucleophiles in the Presence of Protic Acids

Takashi ITO* and Tohru YODEN

Department of Materials Chemistry, Faculty of Engineering, Yokohama National University,
Tokiwadai 156, Hodogaya-ku, Yokohama 240

(Received March 17, 1993)

[Mo(η^5 -MeCp)₂H₂] (3) (MeCp = C₅H₄Me) which was prepared by the reaction of MoCl₅ with Na(MeCp) in the presence of NaBH₄ was treated with *p*-toluenesulfonic acid (TsOH) to give [Mo(η^5 -MeCp)₂H(Ots)]. Reactions of 3 with allyl alcohol in the presence of TsOH yielded [Mo(η^5 -MeCp)₂(η^3 -C₃H₅)]⁺TsO⁻ whereas

the similar reaction with 2-methyl-2-propen-1-ol yielded [Mo(MeCp)₂CH₂CH(Me)CH₂OH]⁺TsO⁻. When the reactions of 3 with protic acid HA (HA = TsOH and AcOH) were carried out in the presence of acetone or 4-*t*-butylcyclohexanone, [Mo(η^5 -MeCp)₂A₂] were formed accompanied by the reduction of ketones to the corresponding alcohols. The results are compared with the similar reactions with cyclopentadienyl analog of 3.

Since the first synthesis of [Mo(Cp)₂H₂] (Cp = η^5 -C₅H₅) (1) by Green et al. in 1961,¹⁾ extensively rich chemistry of 1 has been developed.²⁾ Some of its typical reactions include, for example, the insertion reactions of alkenes or alkynes into its Mo–H bond(s) to give alkyl- or alkynylmolybdenum derivatives,^{3–6)} photochemical cleavage of Mo–H bonds to generate reactive molybdenocene intermediate,^{7,8)} reactions with alkyl halides⁹⁾ or Lewis acids such as Fe(II), Sn(IV), BuLi, and AlR₃,^{10–14)} and the protonation with Brønsted acids.^{1,15)} The trihydridomolybdenum cation derived from the last reaction proved to be an important key compound which shows versatile reactions toward nucleophiles.^{15–19)}

In view of the fact that most reactions of complex 1 depend upon the basic nature of its metal center, it is interesting to see how these reactivities change by introducing the substituents into the cyclopentadienyl ligands in 1.

In spite of ample reports on the [Mo(Cp)₂H₂] chemistry as described above, little has been studied on the chemistry of the type 1 complex with the substituted cyclopentadienyl ligands. This may be partially ascribed to the difficulty in the preparation of such complexes, especially that with the permethylcyclopentadienyl ligand, [Mo(Cp')₂H₂] (Cp' = η^5 -C₅Me₅) (2), as pointed out by Cloke et al.^{20,21)} Although we recognized that 2 can be prepared by the analogous reaction path used for the preparation of 1, the yield is not satisfactory for the purpose of utilizing it as a starting complex for various reactions.²¹⁾

In order to see the substituent effect of Cp ligand on the electrophilic reactions of the complex of the type 1, we prepared the monomethylcyclopentadienyl analog of 1, [Mo(MeCp)₂H₂] (3) (MeCp = η^5 -C₅H₄Me) (Chart 1), and examined some of its reactions with Lewis bases.

Results and Discussion

Complex 3, which is a MeCp analog of 1, was synthesized by following essentially the same procedure as

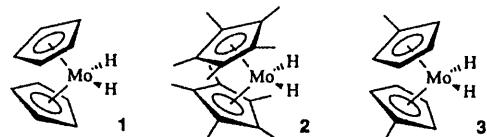
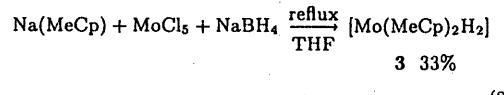
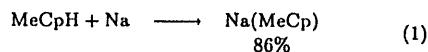


Chart 1.

utilized for the preparation of complex 1. The complex was isolated as yellow needles after sublimation in vacuo in yield of 33% based on MoCl₅ used (Eqs. 1 and 2).



Spectral data for complex 3 are shown in Table 1. In addition to the characteristic $\nu(\text{Mo}-\text{H})$ band at 1855 cm⁻¹, stretching and deformation bands of CH₃ group at 2400–3000 cm⁻¹ and at 1495 and 1450 cm⁻¹, respectively, are observed in the IR spectrum of 3. The $\nu(\text{Mo}-\text{H})$ band at 1855 cm⁻¹ is intermediate between those of [Mo(Cp)₂H₂] 1 (1845 cm⁻¹) and [Mo(Cp')₂H₂] 2 (1860 cm⁻¹), in accordance with the electron-donating nature of the methyl substituent on Cp ring.

In contrast, ¹H NMR chemical shift values of the hydrido hydrogens of complexes 1–3 are not necessarily in the order of electron density of the central metal which is expected from the number of the electron-donating methyl substituent: Thus the chemical shift for 1 ($\delta = -8.76$) is the highest and that for 2 is intermediate value ($\delta = -8.35$). This result may suggest the existence of some other factors such as steric influence which governs the chemical shift of the hydrido hydrogen of this type of complex.

Since unsubstituted dihydride 1 has been known to be protonated with *p*-toluenesulfonic acid (TsOH) in dieth-

<CSJOUR>

<FM><UBER>...<RD>...<VR>199 </VR><RCD>...

<TIC><ALT>Bis(<IT>η: /IT>⁵-methylcyclo...

<SPT>Synthesis and Reactions with Nucleophiles

<ALT>Synthesis and Reactions of [Mo(<IT>η: /IT>...

<AUG><AU AFFRID=A01'><PMM>Takashi /><SMM>...

<ARF ID=A01>Department of ...

<ABS>[Mo(<IT>η: /IT>⁵-MeCp) <INF>2</INF>,...

<BY><SB>...

<PS>Since the first synthesis of [Mo(Cp)....

<SRC><STR>Results and Discussion</STR>...

<P>Complex <B0>3</B0>, which is a MeCp analog of <B0>1...

</CSJOUR>

図3. BCSJのSGML方式の文書テキストの例 (...は省略を示す)

Bull. Chem. Soc. Jpn., Vol. 66, p. 2365-2370(1993)を題材とした。

[Synthesis and Reactions of [Mo(η₅-MeCp)₂H₂]]
Bis(η₅-methylcyclopentadienyl)dihydridomolybdenum(IV).

Synthesis and Reactions with Nucleophiles in the Presence of Protic
Acids

Takashi ITO* and John YODEN

Department of Materials Chemistry, Faculty of Engineering, Yokohama National
University, Tokiwadai 156, Hodogaya-ku, Yokohama 240

(Received 199)

[Mo(η₅-MeCp)₂H₂] (3) (MeCp=C₅H₄Me) which was prepared
by the reaction of MoCl₅ with Na(MeCp) in the presence of NaBH₄ was
treated with *p*-toluenesulfonic acid (TsOH) to give [Mo(η₅-

MeCp)₂H(Ots)]. Reactions of 3 with allyl alcohol in the presence of
TsOH yielded [Mo(η₅-MeCp)₂(η₃-C₃H₅)]⁺TsO⁻ whereas the
similar reaction with 2-methyl-2-propen-1-ol yielded

[Mo(MeCp)₂CH₂CH(Me)CH₂OH]⁺TsO⁻. When the reactions of 3 with
protonic acid HA (HA=TsOH and AcOH) were carried out in the presence
of acetone or 4-*t*-butylcyclohexanone, [Mo(η₅-MeCp)₂A]₂ were
formed accompanied by the reduction of ketones to the corresponding
alcohols. The results are compared with the similar reactions with
cyclopentadienyl analog of 3.

§8. Results and Discussion
Complex <#b>3, which is a MeCp analog of <#b>1...

...
} Since the first synthesis of [Mo(Cp)₂H₂] (1) by Green et al.
in 1961,¹⁾ extensively rich chemistry of 1 has been developed.²⁾ Some of its typical
reactions include, for example, the insertion reactions of alkenes or alkynes into its

図5. BCSJのRTF形式のダンブリストの例 (...は省略を示す)
Bull. Chem. Soc. Jpn., Vol. 66, p. 2365-2370(1993)を題材とした。

図4. BCSJ用スタイルによりMS Wordで作成した電子掲載論文
Bull. Chem. Soc. Jpn., Vol. 66, p. 2365-2370(1993)を題材とした。

表1. Bull. Chem. Soc. Japan電子投稿用スタイル一覧

	スタイル名	内容	書式
フロントマター	Category	分類	行頭字下げなし
	Running title	Running title	14 Point, 中央揃え, 行頭字下げなし
	Title	論文題名	14 Point, 中央揃え, 太字, 行頭字下げなし
	Sub title	副題	中央揃え, 太字, 行頭字下げなし
	Author	著者名	中央揃え, 行頭字下げなし
	Address	研究の行われた場所	中央揃え, 行頭字下げなし
	Received	受理年月日	中央揃え, 行頭字下げなし
	Synopsis	Synopsis	行幅の縮小
本文	Normal	通常の段落	
	Section title	節タイトル	中央揃え, 太字, 行頭字下げなし
	Subsection title	項タイトル	太字, 行頭字下げなし
バックマター	References	引用文献	行頭字下げなし
	Table	Table captions	行頭字下げなし
	Scheme	Scheme captions	行頭字下げなし
	Figure	Figure captions	行頭字下げなし
	Chart	Chart captions	行頭字下げなし
Contents with Illustrations	CI-Title	CI 用タイトル	行頭字下げなし
	CI-Author	CI 用著者名	行頭字下げなし
	CI-Summary	CI 用要約	行頭字下げなし
著者の略歴	Profile	著者の略歴	行頭字下げなし

各スタイルは Normal スタイルを元として設定した。書式欄には Normal スタイルとの相違点のみを示してある。Normalスタイルの書式は、Times 12 point、左寄せ、double space、行頭字下げ、である。

2)文章中に書けるような簡単な式も書いてはいけないのか?、3)表が別扱いなのは不便、などのコメントや質問があった。1)については、投稿者と編集側でフォント指定が異なると変換ミスが起きたため、ギリシャ文字だけでなく化学でよく使用される記号も含めてBCSJ投稿用フォントを作り、投稿者に配付することを考えている。2)は、簡単な式は書いてよいことをユーザ・マニュアルに明記することにした。3)については実現法を検討することとした。また、数式を多用する物理化学分野の投稿者には少数ではあるがLaTeXでの投稿希望者がいることから、実現するよう検討することとした。

5. おわりに

今後もモニターの意見を取り入れ、さらに改良を図っていきたい。

文献

- 1) ISO 8879-1986, Information Processing - Text and Office System - Standard Generalized Markup Language (SGML), Oct. 15, 1986.
- 2) C.F. Goldfarb, The SGML Handbook, Oxford University Press, 1990, 664pp.
- 3) 石塚英弘, SGMLによる情報知識学会誌の編集印刷について, 情報知識学会誌, 1巻, 1号, p. 24(1990)
- 4) 根岸正光, 「SGML実験誌」の出版について, SGML実験誌 1991, pp. i-iii(1991).
- 5a) 伊藤卓, 化学と工業, 日本化学会欧文誌の全文データベース化と電子出版化への移行について, 46巻 1号, pp. 92-95 (1993); 5b) 伊藤卓, SGMLを用いた学術論文誌の電子出版化—日本化学会欧文誌の場合一, 情報の化学と技術, Vol. 43, No. 12, pp. 1089-1096(1993); 5c) 上嶋邦義, 学術論文誌の全文データベース化と電子出版の実際, 医学図書館, Vol. 41, No. 1, pp. 38-45(1994); 5d) 石塚英弘, 伊藤卓, 楊敏明, 千原秀昭, 中西敦男, 田中洋一, 日本化学会欧文誌のSGML形式全文データベースの構築・印刷そして検索, 情報処理学会情報学基礎研究会資料, 29-1 (1993.5).
- 6a) 志田忠正, 伊藤卓, 化学と工業, 47巻, 3号, p. 270 (1994).
- 6b) スタイルファイル利用マニュアル, 凸版印刷(株), 1994.9, 9pp.
- 7) 石塚英弘, 伊藤卓, 千原秀昭, 根岸正光, 中西敦男, 田中洋一, 全文検索システムのリソースとしてのSGML方式データベース, 情報処理学会情報学基礎研究会資料, 33-7 (1994.5).
- 8) H. Ishizuka, "The reception of SGML based electronic publishing by Japanese scientific community", Proc. 47th FID(Int. Fed. Inf. Doc.) Conf. Cong., pp. 505-508 (Oct. 1994, Omiya).