

困ったときに役立つ電子化マニュアルを目指して

○中川 裕志, 森 辰則, 松尾 衛

横浜国立大学工学部

複雑な電子機器やソフトウェアが多くなってきている今日、マニュアルの重要性は高まっている。しかし、一方でマニュアルは依然としてとっつきにくかったり、読みにくかったりする。このような現状を工学的技術に応用することによって改善することは、意義あることである。本報告では、この目的にそって電子化マニュアルを構築するにあたって、(1) 電子化マニュアルの概念、(2) 考慮すべき事項と解決すべき技術的問題点、(3) 我々の研究結果および試作システム、などについて述べる。

Towards Electronic Manuals Helpful In Need

○Hiroshi Nakagawa, Tatsunori Mori, Mamoru Matsuo

Engineering, Yokohama National University

Complex and/or complicated electronic apparatuses, computer systems, and softwares are found everywhere these days. However their manuals are still unfriendly or hard to read for users. This situation could be and should be improved by computer technologies. For this purpose, in this paper, we report (1) the concept of intelligent electronic manual system, (2) technical problems to be solved to implement that kind of system, and (3) our prototype system and the results of our research effort towards problems related to (2).

1 はじめに

ハードは一流、ソフトはまあまあ、マニュアルは joke、という風評が日本の工業製品に寄せられているらしい。聞き流しておくのが、大人の振舞いということかもしれない。しかし、もし実態がそのようであったとすれば、これは確実に日本の工業製品の競争力を奪っている。マニュアルの品質向上の問題の深刻さには軽視できないところがあるのではないだろうか。

マニュアルの問題を考える別の視点として、ソフトウェア、ハードウェアを問わず製品の機能が向上した反面、その機能、操作法ともに複雑の度を増していることが挙げられる。必然的に利用者がマニュアルに頼る割合が高くなる。

以上のような状況から見て、問題の解決法は 1) マニュアルの文章の質の向上、2) 情報処理技術を応用する電子化マニュアル、のふたつが考えられる。1) については、(海保博之, 加藤隆, 堀啓造, 原田悦子, 1987) に詳しく述べられている。その内容は以下のような項目に分けて考えられる。

- a 文の日本語としての質の問題。例えば、「視点を一貫させよ」、「文を短くせよ」など。
- b 談話としての分かりやすさの問題。例えば、「一つの文には一つの操作だけ書け」など。
- c 図や表の使い方の問題。例えば、「実物よりイラストを使え」など。
- d 文書構造の問題。例えば、「目次を階層化せよ」など。

もちろん、そこで述べられている処方箋は非常に有用なものである。また、公にされることは少ないようだが、企業においても、そこで述べられているようなマニュアル作法に関する努力がなされているに違いない。さて、上記 a の問題は言語そのものを相手にするだけに、工学的手法を適用しようとするなら、自然言語処理、それもかなり深いレベルの意味分析技術が必要である。これは単に言語的に精密な分析を要求するだけでなく、多くの場合、そのマニュアルの扱う機器そのもの、あるいは関連する分野に関する固有領域知識を大量に必要とする。自然言語処理の情報リソースとしては(日本電子化辞書研究所, 1995; 国立国語研究所, 1996)があるが、これらは一般的な世界知

識を扱ったリソースであり、上で述べたような固有領域知識に関するリソースではない。さらに、このような固有領域知識は、特に日進月歩の技術分野では絶えず変化しており、情報リソース化に多大の労力を要する。例えば、3年前に Java や Netscape というターミノロジーがどれだけ市民権を得ていたかを考えてみられたい。したがって、現在の自然言語処理技術では a のレベルの計算機処理はかなり困難な問題である。b については、工学的アプローチがより困難である。修辞構造に関する多くの研究、例えば (Mann & Thompson, 1992) があるが、筆者の知る限り、工学的な評価基準を与える一般的方法是提案されていない。c, d については(海保博之他, 1987)では心理的効果を勘案しつつ書き方の上での示唆を与えている。しかし、それらの示唆は電子化されたマニュアルを想定していない。

(海保博之他, 1987)の示唆をマニュアル作りに反映させることが、マニュアルの質の向上に役立つことは疑いないが、それらの示唆はマニュアルライターがいかにマニュアルを書くかという点、すなわち上記の 1) の問題に限定される。2) の電子化されたマニュアルがどのような概念であり、かつどのような技術が現在実用可能か、ないしは開発すべきか、というような問題に答えていく努力がなされねばならない。以下では、電子化マニュアルの概念、必要な技術要素、我々が開発した技術などについて述べる。

2 電子化マニュアルの概念

現在では、ソフトウェアを中心に多くのマニュアルが電子化され、実用に供されている。

UNIX の man

古くから利用されているのは UNIX の man コマンドで検索できるオンラインマニュアルであろう。このシステムは既に名前を知っているコマンドの使い方を調べるときには非常に便利である。しかし、コマンドの名前を知っていなければ使えない。

オンラインヘルプ

Windows95などで使えるオンラインヘルプも、システム使用時には必須の機能いえる。メニューを辿っていく形式であり、かなり使いやすく工夫されている。実際にある機能を実行することも、へ

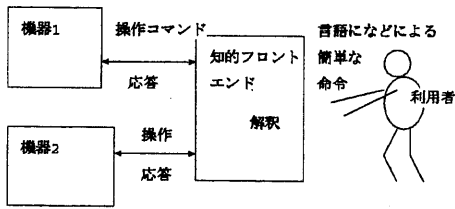


図 1: 知的フロントエンドの概念

ルプの画面からのクリックででき利用者にとってはとても便利である。Windowsもこのオンラインヘルプ無しにはこれほど普及しなかったかもしれない。しかし、このオンラインヘルプは相当な人手をかけて作ったものであろう。

Readme

ほとんど全てのパッケージ・ソフトウェアには俗に readme というファイルが付いている。ソフトウェア利用に先立ってまずこれを読む。使用法なども書かれている。ただし、少なくとも筆者の経験では、readme は始めから最後まで通読することが想定されたような書き方である。

SGML

最近では、以上のような電子化されたマニュアルが普及し、実用に役立っている。一方、将来の電子化マニュアルとしては、SGMLによる規格化が検討されている(笠原, 1996; 藤井義信, 1996)。当面、設計や仕様書の段階から文書をSGML化してデータベースをすることにより、文書情報の共有や再利用を狙っている。また、文書の規格統一による CALS への指向も検討されている。実際のシステムとしては、SGML エディタが開発され、社内利用されている。SGML 化された文書のデータベースの使用法としては、文書管理、内容検索、内容の整合性維持、多組織との互換、などが提案されているが、実用化の進展の度合は明らかでなく、未だ提案段階ではないかと考えられる。

これとは全く次元を異にする方向として、マニュアルに書かれた知識を計算機利用可能な形に変換し、対象の機器の自動運転システムを構築するという流れが考えられる。これは、利用者が対象の機器やシステムを直接操作するのではなく、図1に示すように知的なフロントエンドを介して間接操作する方法と位置付けられる。もちろん、

こうなるとマニュアルも人に読ませるよりは機械に読ませるものというように変わってくるわけだ。しかし、将来、種々の電子機器などが身の回りに溢れたとき、ひとつの操作システム(多分、パソコンを利用するだろう)から、全ての機器を同じインタフェースで操作できることが必要になりそうである。

3 電子化マニュアルに要求される機能

前節で電子化マニュアルの概念について現実的なものから未来的なものまで述べてきた。ここでは、読みやすく、知りたいことが簡単に調べられるような、つまり役に立つ電子化マニュアルを実現するために必要とされる機能をまとめる。

電子ブック化は、マニュアルを読みやすいものにするために必須である。この概念、はハイパーテキスト化によって読者が自由な順序で読んだり、調べたりできるようにしようというものである。また、静止画、動画、音声などもハイパーリンクを張って直接表示、ないしはクリックするだけでアクセスできるようにしようというものである。このような機能は、人手でハイパーテキスト化を行なうなら現在でも実現可能である。ただし、Windows95などと違って利用者数の少ない機器やソフトのマニュアルにもこれだけの手をかけるのは、経済性の点からほぼ不可能であろう。したがって、このような機能を持つ電子ブック化を自動的に行なう必要がある。電子ブックの事例のひとつとしてHMS(島津秀雄, 1996)がある。テキストと画像などからなるページが検索対象であり、目次と呼ばれる階層的インデックスを複数組み合わせ、知りたいことが記述されているページを検索し表示する。このようなシステムの自動構築に当たっては、索引語を抽出し、テキスト中の関連部分と対応付ける作業を自動的に行なわなければならない。索引語と情報検索におけるキーワードとの差は表1にまとめられる。このような索引語を抽出するメカニズムは重要な研究テーマとなる。

電子ブックのようにインデクシングとハイパーテキスト化がなされたマニュアルにおいては、(島津秀雄, 1996)に見られるように、利用者は内容検索機能を使える。内容検索は役に立つ機能だが、考慮すべきは少なくとも次に述べる性質の異なる

表 1: 索引語とキーワードの差異

	索引語	キーワード
検索対象の 文書ベース	単一文書	膨大な 文書群
文書あたりの 数	数 10 から 数 100	数個から 数 10
特徴付ける もの	文書の 一部分	単一の 文書

2 種類の検索に分けられることである。

百科事典的検索 マニュアルに書かれている用語、操作法、概念などを主にキーワードを用いて検索する。例えば、「カナ漢字変換」「フォルダの作り方」など。

トラブル解決法検索 機器やソフトを利用してどうしてよいか分からなくなったとき、あるいは予想した通りの動作をしてくれないとき、に解決策を質問する。

百科事典的検索は情報検索技術に応用すれば比較的簡単なメカニズムで実現できる。(島津秀雄, 1996) の検索もこれに属する。一方、トラブル解決法検索はかなり難しい。1) トラブルの状況説明の解析、2) トラブル原因や解決策の探索、が必要である。両者とも自然言語処理が必要である。すなわち、前者ではトラブル状況説明文の理解が必要である。後者では、マニュアルの記述を利用して解決策を探すなら、マニュアルの内容を理解する処理が必要になる。マニュアルの記述に頼りたくないなら、作成に手間のかかる故障診断エキスパートに逆戻りになる。

この他に、図、写真、ビデオの扱いも重要である。図や写真のような静止画像であっても、単なる表示を越えて、角度や視点を変えて表示させるような技術が望まれる。幸い、VRML など、この目的に適した言語が出回り始めている。動画にも視点の変換などの機能があるとうれしい。これは virtualized reality という概念として研究されているようである。

さて、ここまでで述べた電子化マニュアルの作成は、人手によって行えば労力はかかるものの現

在の技術で十分実現可能である。内容検索にしても、想定される利用者からの質問を徹底的に網羅しておけば、かなり満足のいく応答をする検索システムができそうである。しかし、先に述べたように人手による作成は経済的でないから、多くのマニュアルを人手で電子化することは非現実的であろう。そこで焦点は、前節で述べた意味での電子化マニュアルの自動作成になる。以下の各節で、この自動作成のために必要な技術課題を列挙する。また、これらの課題を解決する技術の現状、およびこれらの問題に対して我々が得た研究成果について触れる。

4 索引語自動抽出

マニュアル文書に利用者を使いやすい構造を与えるためには、1) 章、節などの構造と、2) 質のよい索引語、が必要である。1) については SGML をベースにしてデータベース技術を直接に応用しようという研究(古舘文裕, 1996; 波内みさ, 1996) があるが、利用者が内容に関して知りたい場合は、やはり情報検索的な技術が必要になる。2) に関連しては情報検索の分野でキーワード抽出技術として(Ogawa, Bessho, & Hirose, 1993) など数多くの研究が行なわれてきた。しかし、マニュアルの索引語抽出においては表 1 に示したように若干性質の異なる語を抽出しなければならない。そこで我々は、索引語になるような重要な概念を表すことばには複合語が多いことに着目し、重要度の高い複合語を自動抽出する方法を提案した(中川裕志, 森辰則, 松崎知美, 1996)。この方法は簡単に言えば、1) 多くの複合語の要素となる単名詞ほど重要度が高いこと、2) 複合語の重要度は、その要素である単名詞の重要度の関数として定義される、というふたつのアイデアに基づく。形式的には以下ようになる。まず単名詞に対して前方、後方接続数を提案する。

定義 1 与えられたマニュアルにおいて、単名詞 N の前方接続数は、 $Pre(N)$ と表し、単名詞 N の直前について、 N と複合名詞を作る異なる名詞の数とする。単名詞 N の後方接続数は、 $Post(N)$ と表し、単名詞 N の直後について、 N と複合名詞を作る名詞の数

$Pre(N)$ と $Post(N)$ が単名詞の重要さを表す。複合名詞 $NC = N_1, \dots, N_k$ の重要度 $Imp(NC)$ は基

本的にはこれらの積とする。ただし、 $Pre(N_i)$ と $Post(N_i)$ がひとつでも0になったために全体の重要度が0になることを避けるために次式を採用する。

$$Imp(NC) = \prod_{i=1}^k ((Pre(N_i) + 1) \cdot (Post(N_i) + 1)) \quad (1)$$

これは長い複合名詞ほど重要度が高いという結果を導く。我々はこの他に、 $Pre(N_i)$ と $Post(N_i)$ の総和、あるいはこれらを $1/k$ 乗根、あるいは $1/k$ 倍して複合語の長さを正規化するような関数も考えた。しかし、表2に示した5種類のマニュアルでは結局、上で示した積による関数が1番よい評価を得た。評価は適合率と再現率の逆数の和の逆数(F-measureと呼ばれる。)で行った。また、表3には、これらのマニュアルの索引語候補数、人手で切り出した正解索引語数を記した。

表2: 本研究で用いたマニュアル

マニュアル	総文数	大きさ (kB)
コンピューターソフトウェア (JUMAN)	436	31
〃 (SAX)	433	28
〃 (「たまご」)	628	30
家庭用ビデオデッキ (三菱電機 HV-F93)	1461	69
家庭用ゲーム機 (SONY PlayStation)	131	7

上記の関数の値の大きさの順に並べられた複合語から、索引語を選ぶことになる。この方法も単純に上記関数の値のしきい値を使うだけではうまくいかない。そこで我々は窓方式という方法を提案した(詳細は(中川裕志他, 1996)を参照されたい)。これによれば、抽出した索引語の品質として、上記5種類のマニュアルの平均で再現率70.5%、適合率67.5%という値を得ることができ

表3: 正解索引語数と索引語候補数

マニュアル	正解索引語数	索引語候補数
JUMAN	104	882
SAX	216	797
たまご	102	786
ビデオ	264	1216
ゲーム機	41	150

た。特筆すべきは、この結果は、マニュアル毎に最適化したものではなく、抽出において全マニュアルで同一のパラメータを用いて得られることである。もちろん、マニュアル毎の最適値はもう少しよい。

5 内容検索システム

マニュアルに限らず本というものは、そこに何が書かれているかは読んでみなければ分からない。しかし、知りたいことがどこに書かれているかを知る手がかりはある。つまり、目次と索引である。(長尾, 1994)では目次の持つ情報について分析し、文献検索に役立つことを説明している。また、目次で表わされる文献ようなもの構造を利用した検索の研究もなされている(Fuller, Mackie, Sacks-Davis, & Wilkinson, 1993)。一方、索引を利用する方法としては黒橋の読書支援(黒橋禎夫, 白木伸征, 長尾真, 1996)の研究がある。これは索引語の本文中での分布を調べ、高密度で分布する部分を選ぶと、その索引語を定義している部分が得られるという方法である。黒橋の方法はマニュアルの内容検索にも応用できると考えられるが、単独の索引語を検索のキーにしているため、複数の質問キーワードを持つ質問、例えば「変換辞書を新規登録するときに変更が必要なファイルは何か?」などには直接答えられない。我々は、情報検索の技術を応用して以下に述べるようなシステムを試作している。

まず、マニュアルをセグメントと呼ぶ検索対象となる単位に分割する。セグメントの決め方はいくつかが方法がある。例えば、一定字数で切断してセグメント化する方法もある。しかし、ここではマニュアルの書き手の意図が明確な形式的な形

で現れる章、あるいは節という構造的単位のうちの最小のものを選ぶことにした。つまり、目次に現れる最も下位の単位を検索および表示の単位として使う。このようにセグメント分割したマニュアルを図 2 に示すような流れで検索する。

質問文に見合うセグメントを検索する方法としては、いくつかの方法を比較してみた。その結果、前記のマニュアルのうち JUMAN、SAX、ビデオデッキ、「たまご」において、1) 単純名詞の tf.idf の重みを利用するベクトル空間法と 2) 複合語をひとつの単語と見なしての tf.idf 重みを利用するベクトル空間法、というふたつの方法でよい結果が得られた。再現率、適合率を図 3 に示す。ベクトル空間法による質問とセグメントの一致の高い順にセグメントを並べ、そのうち上位から何セグメントまでを採用したかを、図の横軸が示す。縦軸は再現率と適合率である。

ここで問題は最もランキングの高い、つまり質問に関連性の高いセグメントを見た場合、再現率、適合率とも 30% から 40% 程度という結果しか得られないことである。つまり、質問者は欲しいセグメントをすぐに取り出せない。もっとも正統な方法は検索エンジンを工夫して適合率、再現率を向上させることであり、我々もこの方向での努力を続けてはいる。ただし、即応的な方法として、セグメント表示方法を工夫することも検討しなくてはならない。我々は、検索されたセグメントをその先頭からではなく、質問中の単語の現れた箇所付近から表示する方法を採用している。検索結果の具体例を図 4 に示す。

この方向での研究として、(Fuller et al., 1993) がある。そこでは、セグメント単位で文書を検索する passage retrieval の考え方を利用し、検索文書の持つハイパーテキストとしての構造とセグメント内容を同時に表示するようなシステムを提案している。文書構造の表示により必要なセグメントの所在を質問者に分かりやすくしようというものである。

6 将来的課題

前節で述べた内容検索できるマニュアルの進化するべき方向について述べる。まず、マニュアルをハイパーテキスト化するにあたって、張るべきリンクだが、通常のハイパーテキストで用いられる

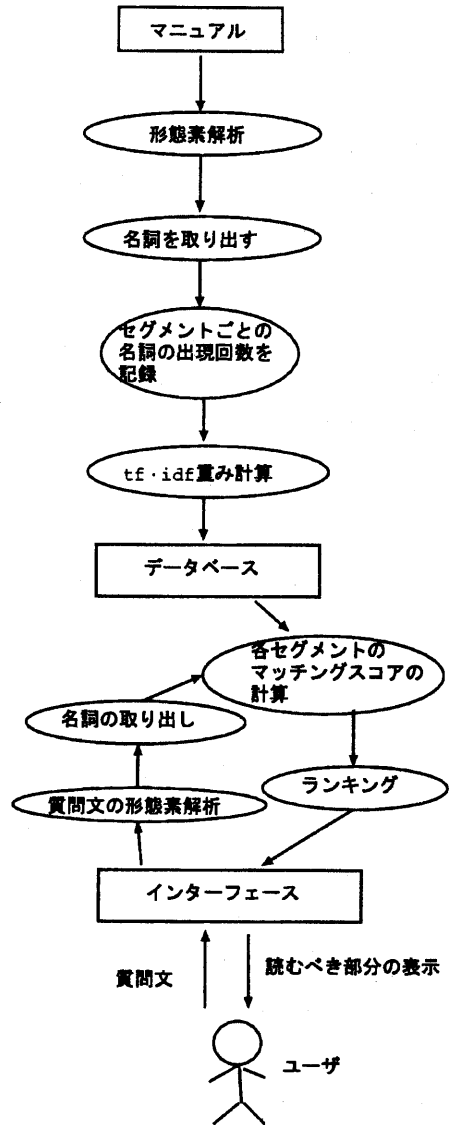


図 2: 内容検索システムの概念

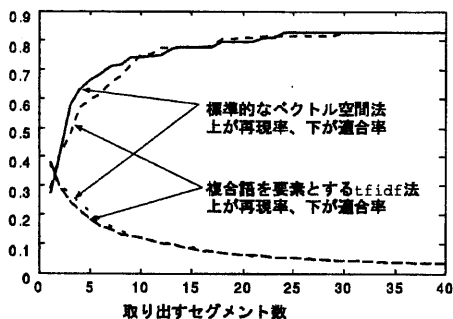


図 3: 内容検索の再現率と適合率

章や節、図などのリンクの他に、次の 2 種類のも
のが有力である。

- 定義部分と参照表現をつなぐリンク。用語の
定義を忘れたとき役立つ。
- 複数のマニュアルの関連部分に対応づけるリン
ク。例えば、初心者マニュアルに書かれて
いる事項について、より詳しく知りたいとき
リファレンスマニュアルの対応部分を調べる
が、このような場合に役立つ。

このようなリンクを自動的に検出し、張っていく
方法については現在検討中であり、近く発表す
る。より高度な内容検索として、利用者がトラ
ブルに遭遇したとき、その状態を入力すると、対応
する処置を表示してくれるという機能がある。状
況と対応処置の関係は人手でつけることができ
るが、手間のかかる作業である。この対応を自動
的に発見できることが望ましいが、かなり高度な
自然言語処理技術が必要である。我々は、その手
始めに、マニュアルの文において省略された主語
を求める方法を開発した (Mori & Nakagawa, 1995,
1996)。これにより、操作の記述があったとき、誰
が行なうべき操作かが決まる。これに、「利用者
がトラブル状態にあるとき、行なうべき操作で
あること」を示す言語表現を適用すれば質問文に
対応する困った状況への対応処置が記述されて
いる場所を検索できるのではないかと予想され
る。これについても検討中であり、結果がま
とまり次第報告したい。

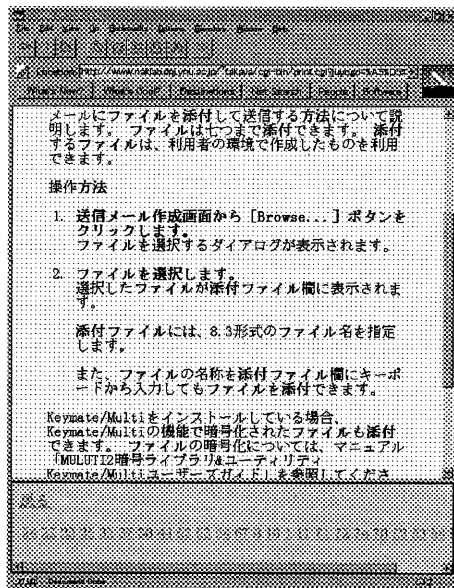


図 4: 検索結果画面例 (日立製作所より提供いた
だいたマニュアル「Group max WWW ユーザーズ
ガイド」による)

7 まとめ

内容検索できる電子化マニュアルについて、そ
の概念、技術の現状、我々が開発しているシス
テムの概要、などについてまとめた。このよう
な研究は、大きな流れで言えば、いわゆる情報
過負荷状態から人々を救おうというシステムの
研究である。マニュアルにおいては、過負荷
状態は大きな問題であり、産業界とアカデ
ミックサークルが強力して、このような研究
の推進を計らなければならない時期にきて
いる。

謝辞

本研究はIPA「創造的ソフトウェアプロジェクト」
の援助を受けている。また、日立製作所から
マニュアルの提供を頂いている。

参考文献

Fuller, M., Mackie, E., Sacks-Davis, R., &
Wilkinson, R. (1993). Structured Answers

- for a Large Structured Document Collection. In *ACM-SIGIR '93*, pp. 204-213.
- Mann, W. C. & Thompson, S. A. (1992). Rhetorical Structure Theory and Text Analysis. In *Discourse Description*, pp. 39 - 78.
- Mori, T. & Nakagawa, H. (1995). A pragmatic approach to zero pronoun resolution in Japanese manual sentences. In *The Proceedings of NLPRS'95*, pp. 296 - 301.
- Mori, T. & Nakagawa, H. (1996). Zero Pronouns and Conditionals in Japanese Instruction Manuals. In *The Proceedings of COLING'96*, pp. 782 - 787.
- Ogawa, Y., Bessho, A., & Hirose, M. (1993). Simple Word Strings as Compound Keywords: An Index and Ranking Method for Japanese Texts. In *ACM-SIGIR'93*, pp. 227-236.
- 海保博之, 加藤隆, 堀啓造, 原田悦子 (1987). ユーザ・読み手の心をつかむマニュアルの書き方. 共立出版.
- 笠原健成 (1996). “SGML適用による電子化の実践.” 情報処理学会 DD 研究会 96-DD-2, pp. 1-8.
- 黒橋禎夫, 白木伸征, 長尾真 (1996). “出現頻度分布を用いた語の重要説明箇所の特定.” 情報処理学会研究報告 96-NL-115, pp. 43-50.
- 国立国語研究所 (1996). 新分類語彙表. 国立国語研究所.
- 島津秀雄 (1996). “事例ベース検索のツール.” 農林水産業の高度情報システム: 農林水産技術協会刊, pp. 139 - 143.
- 中川裕志, 森辰則, 松崎知美 (1996). “日本語マニュアル文における名詞間の接続情報を用いたハイパーテキスト化のための索引語の抽出.” 情報処理学会研究会資料 96-NL-116, pp. 65-72.
- 長尾真 (1994). 電子図書館. 岩波書店.
- 波内みさ (1996). “OODBによるSGML文書データベースの設計.” 情報処理学会データベースシステム研究会 96-DBS-109, pp. 311-316.
- 日本電子化辞書研究所 (1995). EDR 電子化辞書仕様説明書. 日本電子化辞書研究所.
- 藤井義信 et al (1996). “ソフトウェア設計情報共有化の為にソフトウェア CALS システムに関する一考察.” 情報処理学会 DD 研究会 96-DD-2, pp. 9-16.
- 古館丈裕 et al (1996). “構造化文書データベースに対するラッピング手法の提案.” 情報処理学会データベースシステム研究会 96-DBS-109, pp. 305-310.