社団法人　情報処理学会　研究報告
IPSJ SIG Technical Reports

2005－ＤＤ－51（４）
2005／7／22

デジタルドキュメント研究に関する傾向についての続報
－デジタルドキュメント10年の傾向－

三田 虎史† 秋元 良仁† 斎藤 伸雄†

†凸版印刷株式会社　情報ビジネス開発本部　研究開発部

〒112-8531 東京都文京区水道 1-3-3

E-mail: †takeshi.mita@toppan.co.jp

あらまし　本稿は，近年の情報処理技術におけるデジタルドキュメント研究の傾向について，情報処理学会デジタルドキュメント研究会の研究報告を元に分析・報告を行った「デジタルドキュメント研究 10 年の傾向」[1]の続報である．

オフィス文書など，紙媒体の電子化に端を発したデジタルドキュメント研究は，情報のデジタル化に大きく貢献してきた．デジタルドキュメント研究の 10 年の傾向を分析する事は，今後のデジタルドキュメント研究を行う上での指針となりうる．そこで，デジタルドキュメント研究会における 252 件の研究報告をテキストマイニングツールにより分析し，歴史的なトピックと比較しながら考察を加えた．XML や携帯電話，Semantic Web などのトピックスとの関係がみられたので，それらの傾向について報告する．

キーワード　デジタルドキュメント，テキストマイニングツール，XML，携帯電話，Semantic Web

Continued report on tendency of digital document research
－Tendency of digital document research on 10 years－

Takeshi Mita† Ryoji Akimoto† and Nobuo Saito†

† Research and Development Dept. Info-communication Business Debision, Head Office, Toppan Printing Co., Ltd.

1-3-3 Suido, Bunkyo-ku, Tokyo, 112-8532 Japan

E-mail: †takeshi.mita@toppan.co.jp

Abstract　This paper is a continued report of the "Tecdency of digital doument research 10 years"[1] that analyzes and reports based on the reports of research in SigDD (Special Interest Group on Digital Documents).

The digital document research, which started by computerization of the paper medium such as office documents, has greatly contributed to the digitalization of information. It is useful for the digital document research in the future to analyze the tendency of the digital document research on ten years recently. Then, 252 reports of research in SigDD are analyzed with the text-mining tool, and consideration is added while comparing it with a historical topic. We report tendency of the relation to topics such as XML, mobile phone, and Semantic Web.

Keyword　Digital document, Text-mining tool, XML, Mobile phone, Semantic Web

1. はじめに

2005 年に 10 年目を迎えた情報処理学会デジタルドキュメント研究会では，デジタルドキュメントを中心とする，情報処理の各分野の横断的な研究活動を行っている．研究対象分野は，文書記述，作成技術，管理技術，表示技術，情報流通・活用技術，基礎技術，応用技術など広範囲に及んでいる．

我々は，デジタルドキュメント研究会の研究報告の傾向を読み解く事で，近年のデジタルドキュメントや文書情報を対象とした研究傾向を把握し，今後の研究への指針を導き出せるのではないかと考えた．そこで，過去の研究報告について，トピックの抽出を行い，更にそのトピックを元に経年の傾向を分析し，デジタルドキュメントを取り巻く環境変遷の傾向を見た．

以下，2 章では分析手法を解説し，3 章から 5 章では各分析結果を，6 章では分析に基づいたまとめを述べる．

2. 分析手法

2.1 分析対象

1996年6月から2005年1月までの，デジタルドキュメント研究会における研究報告252件の抄録情報，および本文情報を分析の対象とした．各書誌情報については，情報処理学会電子図書館[2]より取得した．

2.2 テキストマイニングツール

研究報告の傾向を分析するにあたり，「デジタルドキュメント研究10年の傾向」で分析ツールとして利用した，TRUE TELLER[3]を用いた．TRUE TELLERは，テキスト情報を解析するテキストマイニングツールであり，主に顧客のアンケート情報の解析によるマーケティング分析などに活用されている．分析対象のテキストデータを形態素解析し，出現単語による傾向分析を行う事ができる．

テキストデータ内の単語出現頻度を示す「単語ランキング」や，単語同士の関係を主成分分析などに基づいて2次元にプロットする「マッピング」機能がある．また，年代や特定単語を含むなどの属性を指定してグループを形成し，グループごとに「グループ分析」が行える．グループは，テキスト内の単語の出現傾向から，自動的に生成する事も可能である．また，「キーワード抽出」機能により，全体と比べて，当該グループで特に出現頻度の高い単語を取り出す事ができる．

2.3 引用分析

引用分析とは，論文や研究報告など文献間の引用・被引用の関係を明らかにする事で，特定主題に関する文献の引用状況の調査を行い，研究動向の分析や雑誌の利用度分析などに活用するものである．

実際の引用関係では，報告書内の参考文献の記述により，どの文献を引用しているかを確認することは容易だが，一方でどの文献で引用されているかを調査することは困難である．そこで本研究では，国立情報学研究所による学術論文データベースである「CiNii（Citation. Information by NII：サイニイ）」[4]での検索結果を一つの指標とした．また，デジタルドキュメント研究会の報告間での引用関係について，各報告内の参考文献の項からカウントし，比較対照とした．

2.4 トピックワード出現時期

一般的なトピックを示す単語（トピックワード）が研究会内で初めて登場した時期を知るために，トピックワードが抄録に登場した時期を時系列にプロットした．各トピックワードについては，各報告内で取り上げられている単語の中で重要だと思われるものを，2.2および2.3節の分析に基づき，筆者が任意に選択した．

近年の情報処理の動きとの比較や，新しいトピックへの取り組み傾向の把握を意図している．

3. テキストマイニングツール分析

3.1 全体分析

テキストマイニングツールによって抄録と本文に記載された単語の出現頻度について，10年間を前・中・後の3つの年代に分け，傾向を分析した．

最初に，抄録における頻出単語の傾向を図1および表1に示す．年代別傾向を見ると，デジタルドキュメント研究の対象が，「文書」そのものが話題の中心だった初期に対し，徐々に，「データ」としてより広義の「情報」を取り扱うようになったことが伺える．また，初期に頻度の高かった「SGML」が中期以降見られなくなり，次第に「XML」や「Web」といった単語の頻度が高まっている事が分かる．

一方，本文の分析では，抄録よりも詳細な技術用語の傾向検出を狙ったが，実際には「必要」，「対象」，「場合」などの一般的な単語が上位となり，技術用語が埋没する事となった（図2，表2）．また，「情報」，「システム」などの単語は，多くの本文で使用されているため，各報告間の差異が少なくなり，傾向把握に有効な結果が得られなかった．

そのため，以降のテキストマイニングツール分析は抄録文を元に行った．

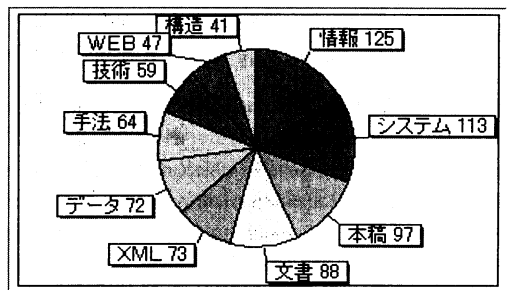

図1　抄録頻出単語（全体）

表1　年代別抄録頻出単語傾向

全体		1996-1998		1999-2001		2002-2005	
単語	頻度	単語	頻度	単語	頻度	単語	頻度
情報	125	文書	30	XML	38	情報	63
システム	113	情報	26	情報	36	システム	55
本稿	97	システム	25	本稿	36	データ	38
文書	88	本稿	24	システム	33	本稿	37
XML	73	SGML	15	文書	25	手法	35
データ	72	手法	14	データ	21	技術	33
手法	64	データ	13	処理	16	文書	33
技術	59	ユーザ	12	今後	15	XML	31
WEB	47	技術	12	手法	15	WEB	29
構造	41	検索	12	技術	14	本研究	23

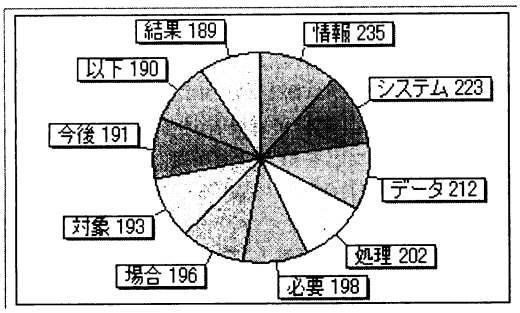

図 2　本文頻出単語

表 2　年代別本文頻出単語傾向

全体		1996-1998		1999-2001		2002-2005	
単語	頻度	単語	頻度	単語	頻度	単語	頻度
情報	235	システム	54	情報	75	情報	109
システム	223	情報	54	システム	67	システム	102
データ	212	対象	51	データ	65	データ	97
処理	202	データ	50	技術	63	今後	91
必要	198	必要	50	処理	62	処理	91
場合	196	以下	49	必要	62	結果	90
対象	193	処理	49	文書	61	場合	89
今後	191	場合	49	他	60	複数	87
以下	190	部分	48	対象	60	必要	86
結果	189	他	47	例	60	中	83

3.2 キーワード分析

単語の出現頻度を見た全体分析では，年代別に推移する話題の傾向が把握できた．

本節では，各年代に特徴的な話題のピックアップを試みた．年代ごとに研究報告を分け，出現に偏りのある単語をキーワードとして抽出した（表 3）．表中のスコアは，χ 二乗値を利用して，± 1 の範囲に収まるように算出されている．他の年代での出現頻度が低く，ある年代に限ってよく出現する単語は，特徴的な単語としてスコアが高くなる．各年代について上位 10 単語をまとめた．

初期の特徴的トピックとして，SGML が挙げられる．90 年代後期は中心的な研究テーマであったが，以降ではあまり話題にのぼらなくなり，技術の中心が XML へと変化していったと考えられる．また，初期のデジタルドキュメント研究会においては，文書自体の電子化や，その際の構造定義に関する研究が多かった．対象となる文書は，マニュアルや新聞が取り上げられていた事がわかる．

中期では，SGML に変わり XML が中心的トピックとなっている．また，機関という単語についてそれを含む報告本文を見ると，医療・教育・金融などの各機関への応用を対象とした研究が行われている．文書自体の電子化から，徐々にその応用へと研究が発展していく様子が見てとれる．

表 3　年代別キーワード傾向

1996-1998		1999-2001		2002-2005	
単語	スコア	単語	スコア	単語	スコア
SGML	0.1365	XML	0.0970	本研究	0.0565
文字	0.0505	最近	0.0319	提案	0.0410
マニュアル	0.0472	考察	0.0262	電話	0.0374
DTD	0.0398	分析	0.0262	メタ	0.0374
操作	0.0373	ドキュメント	0.0245	実装	0.0346
文書	0.0313	機関	0.0236	近年	0.0345
電子化	0.0291	今後	0.0233	携帯	0.0282
調査	0.0269	中心	0.0211	端末	0.0273
現在	0.0247	課題	0.0206	モバイル	0.0267
新聞	0.0245	目的	0.0194	自然言語処理	0.0221

後期では，携帯電話・モバイル端末を対象とした研究が多く取り組まれるようになった．これは，ドキュメントや情報のインフラとして，携帯電話に代表されるモバイル端末が急速に拡大した状況を反映していると考えられる．また，電子化が進んだ結果，データの利活用を目的としたメタデータの研究や，実装についての研究が盛んになってきている．

3.3 主題分析

全研究報告について，各報告内の単語の出現傾向から，類似している報告群のグループを自動生成した．分類するグループ数については，「システム系」，「処理系」，「データ・情報系」，「その他」などの分類を想定し，4 つのグループへの分類を指定した．その結果，表 4 に示す 4 グループに分類された．

表 4　主題別分類結果

分類	報告数
本稿・検索・手法	56
XML・本稿・言語	48
システム・環境・マルチメディア	75
種類・可能性・国際	69

分類を示す単語は，グループで出現頻度の高い 3 語を示す．グループ間およびそこに含まれる単語の関係を把握するために，分類結果を 2 次元にマッピングした（図 3）．プロットの軸には，共起関係（単語の組合せ出現）から求められた第 1 軸と第 2 軸を用いている．図中の二重丸は，分類されたグループの中心を示し，分類や単語間を結ぶ線は共起関係を示している．単語のプロットの大きさは出現頻度に比例している．

図 3 の中心に位置する「XML」を含むグループについて分析する．分類内の報告数は 48 と最少だが，他のグループの中心に位置し，各グループと関連があることが分かる．この事から「XML」を用いた多様な分野の報告があったと考えられる．また，単語の関係から，「文書」や「データ」の「構造」としての「XML」の

活用と，「Web」や「プログラム」に関わる「技術」としての「XML」の活用など，幅広い範囲に「XML」が浸透している事が分かる．

上部に位置する「検索」を含むグループからは，様々な「情報」を「対象」とした「検索」「手法」が提案されている事が分かる．特にデジタルドキュメント研究会では「文書」を「対象」とした「検索」が多く取り上げられている．

右下の「システム」を含むグループでは，「映像」を含んだ「マルチメディア」などに関わる「システム」や「環境」に関する報告があったことが分かる．「XML」を含むグループと「システム」が結びついている事から，「XML」を用いた「システム」が取り上げられている事が分かる．

左下のグループについては，他のグループに入りきらなかったものが包括されている傾向にある．その中でも「国際」的に標準となっている「JAVA」など「処理系」の「技術」の「可能性」が研究されている事が分かる．また「XML」を含むグループとの距離が近い事から，「XML」と強い関連性がある事が伺える．

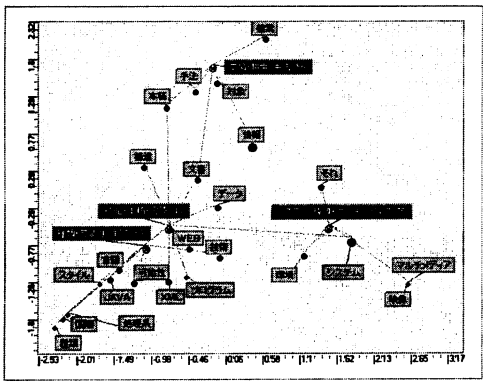

図 3　主題分析マッピング

3.4 グループ分析

抄録全体の単語傾向や主題分析から，トピックとなるキーワードを筆者が選択し，キーワードを含むグループ分けを行い，その中での傾向を探った．調査したグループの分類は以下のとおりである．

- SGML/XML
- インフラストラクチャ
- コンテンツ
- ビジネス関連

3.4.1 SGML/XML

デジタルドキュメントに大きなインパクトを与えた「XML」について，「SGML」と共に調査した(図 4)．

文書の論理構造，意味構造を記述する言語として，CALS[5]への応用などを取り上げられてきた SGML が，1998 年を境に急速に減少している．これは，1998 年 2 月に XML1.0 が W3C 勧告となり，1998 年 5 月には JIS 化されたことが大きく影響している．以降，SGML は XML に取って代わられる．1999 年の SGML に関する 4 件の報告は，全て「SGML / XML」として報告され，XML を解説するために SGML が用いられており，SGML そのものに関する報告は 1998 年以来現れない．

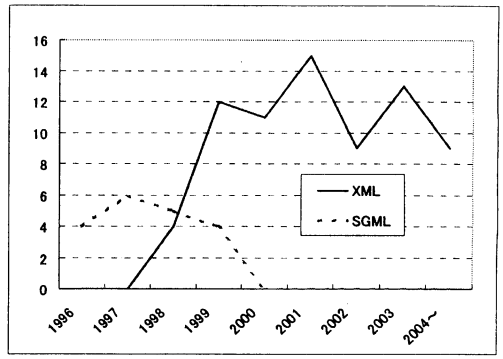

図 4　SGML/XML についての報告数

W3C での勧告直後に研究報告で取り上げられ，デジタルドキュメント研究においての中心となった XML は，それ以降恒常的に取り上げられるトピックとなる．XML 登場初期の 2000 年前後は，何にでも使える魔法の道具として世間一般でもてはやされた時期といえる．研究会では，企業を中心に XML をキーワードとした研究・製品の報告が行われている．1998-2000 年では，XML に関する発表 27 件全てが，企業による発表である（1 件の産学共同を含む）．いかに企業が，積極的に XML に取り組んでいたかが分かる．

2002 年頃より，情報記述手段として XML が一般的となり，その活用手段へと研究対象が変化していく．技術的トピックスにおいては，XLink, XSLT, XQuery, XPath などを使った管理・配信技術が数多く取り上げられている．XSLT を使った変換技術は，2001 年以降にコンテンツのマルチユースなどを目的として研究発表されている．

また，適用分野を限定して XML を活用した標準に関する発表も見られる．旅行分野に特化した TravelXML, 放送用の BML, 地理情報に対する G-XML, 数式用の MathML などが挙げられる．

「XML」を含む報告内の頻出単語とマッピング結果を図5，6に示す．XMLに関する研究報告が，広範囲の分野に及んでいる事がわかる．「文書」「構造」に始まり，XML「データベース」などでの「データ」「処理」や，「WEB」「技術」・「情報」「システム」への活用が行われている．

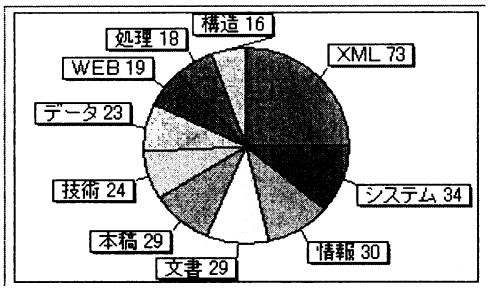

図5　XML関連報告の単語頻度

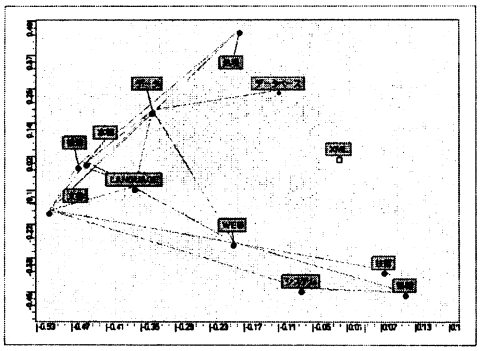

図6　XML関連単語マッピング

3.4.2 インフラストラクチャ

1990年代初期に始まったインターネットは，Webブラウザである Mosaic の登場を境に，1993年頃から爆発的に普及した．インターネットへの接続ホスト数は，デジタルドキュメント研究会が発足した 1996 年では1000万台だったが，2000年代には世界で1億台を突破している．

情報インフラとしてのインターネット，あるいはその上に組織される WWW の発展は，デジタルドキュメントに大きな影響を与え，マルチメディア化，アーカイブ化を急激に進展させ，インターネットを通じた情報の検索，交換，閲覧，加工を可能とし，関連研究が活発に行われた．

そこで，インターネット関連（「インターネット」，「WEB」，「WWW」）および，近年急激に普及を遂げたモバイル関連（「携帯電話」，「PDA」，「モバイル」）を含む研究報告の年別報告数を図7に示す．

社会情勢を反映し，インターネット関連の研究報告は年々増加している．2002年以降では，ブロードバンドを前提とした大量データの処理や，ユーザの利用が盛んな P2P や掲示板に関する研究が発表された．今後は，情報発信・共有の場として急激に利用が拡大している Blog や SNS 関連の研究も行われるのではないかと予測される．

インターネットの普及により，様々な関連技術が研究されているが，その中で注目される動きとして，Tim Berners-Lee によって提唱された Semantic Web[6]がある．デジタルドキュメント研究会でのインターネット関連の報告内に出現する単語のマッピング結果（図8）からも，「今後」の課題として Semantic Web に対する研究が行われていることが分かる．研究会では，2001年より関連報告が増えつつあるが，今後も注目されるテーマであると考えられる．

一方で，インターネットに追従するように，モバイル関連の研究報告も 2001 年以降増加している．XSLT などによるデータの変換や表示に関する研究報告が多いが，よりサービス寄りの研究が進んでいく事が期待される．

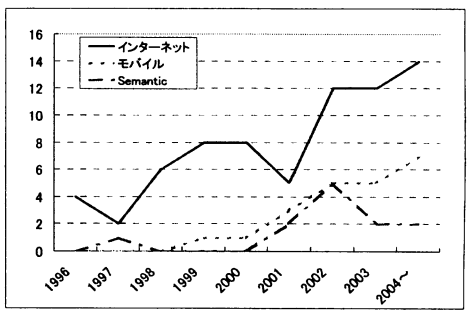

図7　インターネット/携帯関連の報告数

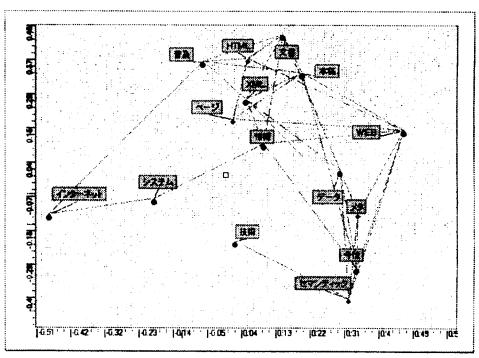

図8　インターネット関連用語マッピング

3.4.3 コンテンツ

電子化の対象は，当初，テキスト情報中心の「文書」が大きく取り扱われていた．ネットワークの整備やPC性能の向上に伴い，音声や画像情報もデジタル化の対象となってきた．これらのデジタルデータを指し示す用語として「コンテンツ」が登場し，現在では一般的に使用されている．そこで，研究報告での「コンテンツ」に関する傾向を見た（図9）．

デジタルコンテンツの普及に伴い，関連研究も増加しているが，連動する動きに「メタデータ」があり，若干遅れながら同期している．

メタデータは，コンテンツの内容を記述する手段として，コンテンツ管理や配信に不可欠であり，コンテンツの増加に伴い，重要性が増していったと考えられる．2003年以降では，RDF[7]やRSS[8]を用いた研究報告が発表されている．

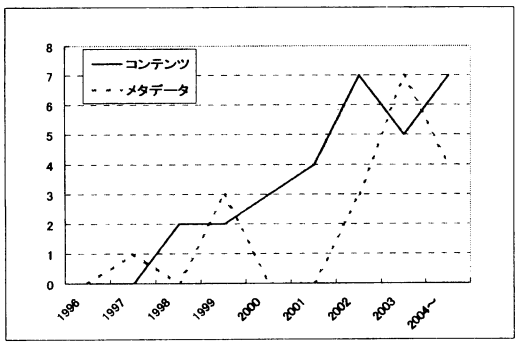

図9　コンテンツ関連の報告数

3.4.4 ビジネス関連

研究報告を俯瞰すると，ビジネス，EC，EDIなど，ビジネス関連の用語を含む研究報告が目に付く．研究会の初期から，CALSや製品マニュアルに関する研究報告が発表され，ビジネスをにらんだ研究が行われていたことが分かる．そこで，ビジネス関連の用語（「ビジネス」，「EC」，「EDI」）を含むグループを分析し，傾向を調査した．

年別の報告傾向（図10）を見ると，2000年前後にビジネス関連の報告が増加している．ビジネス関連の研究報告の多くがXMLを取り扱っている事から(図11)，ビジネスに対してXMLが大きく影響していると考えられる．

2000年前後から活発に研究されるようになったXMLは，帳票などのビジネス関連文書に多く活用された．実際に，E-ビジネスに関わるB2B国際標準においては，XMLベースのものが主流となっている．1999年11月からebXML（2001.5公開）の標準化活動が，2000年8月よりWebサービスの核となるSOAPとUDDI（2000.8公開）の標準化活動が開始されている．研究報告では，ECやEDIについての発表が同時期に挙がっている．

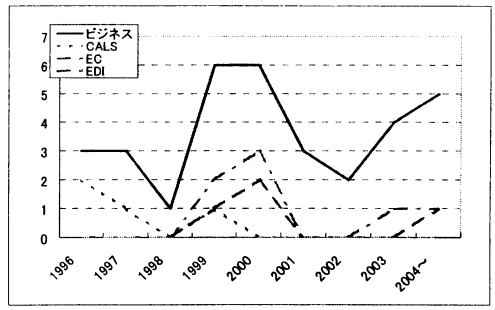

図10　ビジネス関連の報告数

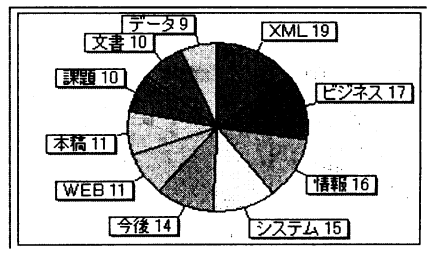

図11　ビジネス関連報告の頻出単語

4. 引用分析

CiNiiと各報告からカウントした研究報告全体の被引用傾向を表5に示す．また，被引用回数が比較的多いものについて，表6にタイトルを列挙する．

表5　引用傾向

	被引用文献数	被引用回数
CiNii	18	28
デジタルドキュメント研究会内	50	83
全体	61	111

CiNiiでの検索結果では，引用される文献数・回数共に少ない．ただし，表6に見られるように，デジタルドキュメント研究会内の引用関係が反映されていない部分が見られた．

研究会内では252件中50の報告が引用先として挙がっている．当然，新しい報告については，引用対象となりにくいことを考えると，研究会内では継続的な

表 6　被引用文献

発行	タイトル	引用された回数			著者
		CiNii	DD研内	合計	
1999.5	オブジェクト指向スクリプト言語Rubyによる XMLの処理	1	7	8	吉田正人／大野邦夫
1999.1	オブジェクト指向スクリプト言語Rubyによる XML応用システムの検討	1	4	5	吉田正人／大野邦夫／藤田大作／前芳久／廣瀬貴幸
1999.7	単語の連想関係によるテキストマイニング	5	0	5	渡部勇／三末和男
1999.7	テキストマイニングのための連想関係の可視化技術	4	0	4	三末和男／渡部勇
1997.5	ORDBによるマルチメディア・ドキュメントの管理	0	3	3	大野邦夫／佐藤和也
1999.3	XML文書のスタイルシート生成方式	0	3	3	森口 修／今村誠／鈴木克志
2001.3	XML文書ワークフロー構築支援方式 － インターネットを用いた設計支援システムにおけるXML文書設計支援方式	0	3	3	今村誠／森口修／鈴木克志
2002.3	モバイル・インターネット環境構築支援システムの検討	0	3	3	大野邦夫／前芳久／吉田正人／近藤治
2002.3	SVG-DOMによるアニメーションと XHTML中心複合文書の可能性	0	3	3	大坂哲司／野村直之
2002.5	セマンティックWebの課題と携帯電話から見た可能性	0	3	3	大野邦夫
2000.5	FIPAエージェントにおけるXMLの適用動向	1	2	3	大野邦夫
2001.7	電子申請におけるXML文書内容検証方式 － 複数XML文書の内容間制約を記述する文書規約記述言語DRDL	1	2	3	今村誠／鈴木克志
2000.7	ハイパーリンクとアンカーテキストを利用した情報検索とランキングの一手法	2	1	3	風間一洋／原田昌紀／佐藤進也

研究が盛んに行われていると考えてよいであろう．

表 6 中の引用された報告を見ると，1999 年～2001 年にかけて，XML 関連の研究が非常に多い事が分かる．タイトルに「XML」がない報告でも関連の深いものが見受けられる．「ORDB によるマルチメディア・ドキュメントの管理（1997.5）」報告は，SGML を使ったオブジェクト指向データベースに関する報告である．また，「SVG-DOM によるアニメーションと XHTML 中心複合文書の可能性（2002.3）」報告は，XSL&XSLT による XML 文書の変換を取り扱っている．デジタルドキュメントにおいて，処理系などに XML が継続的に研究されている事が分かる．

XML 以外の報告では，モバイルや Semantic Web を対象とした「モバイル・インターネット環境構築支援システムの検討（2002.3）」と「セマンティック Web の課題と携帯電話から見た可能性（2002.5）」がある．どちらも 2002 年以降の論文であり，今後主流となりうるトピックだと考えられる．

一方，CiNii の検索結果では，デジタルドキュメント研究会内とは異なった傾向が出た．引用回数が多かった報告は，「単語の連想関係によるテキストマイニング」と「テキストマイニングのための連想関係の可視化技術」であり，外部からの引用である．共に，言語解析に基づくテキストマイニング技術についての研究報告である．言語解析関連の研究は，他にもデジタルドキュメント研究会で発表されているが，情報学基礎（FI）や自然言語処理（NL）など，より専門的な研究会が主流だと考えられる．そのため，外部の研究会との引用関係が強かったと考えられる．

また，表 6 に挙げた被引用回数の多い報告は，全てが企業所属の発表者によるものであった．

5. トピックワードの出現時期

トピックワードについて，デジタルドキュメント研究会の抄録に初めて登場した時期をプロットした年表を図 12 に示す．

トピックワードで灰色のものは，学校所属の発表者による報告からの抽出を示す．二重線で囲まれたトピックワードは，企業所属の発表者の報告からの抽出である．同時に発表があったものについては，二重線に囲まれた灰色で示す．

XML 関連については，W3C の動きと合せて素早い報告が見られる．XML の登場は W3C の勧告とほぼ同調しており，関連技術も XLink（2001.6 Rec.）や XQuery（WD）など，勧告前から積極的に研究されている．XSLT（1999.11 Rec.）や XPath（1999.11 Rec.）の抄録における登場は遅いが，報告本文を調査すると，XSLT が 1999 年 9 月，XPath が 2000 年 3 月に登場している．デジタルドキュメント研究の分野で，XML 関連の研究が積極的に取り組まれてきた歴史が分かる．

2000 年には E-ビジネス系のビジネス用語が多く登場するようになる．これは，ネットワークの拡大により新たなサービスやシステムが生まれ，これらを説明する際に分かりやすい単語が必要とされた事が背景としてあるだろう．また，Web サービスや ebXML などの動きも影響を及ぼしていると考えられる．

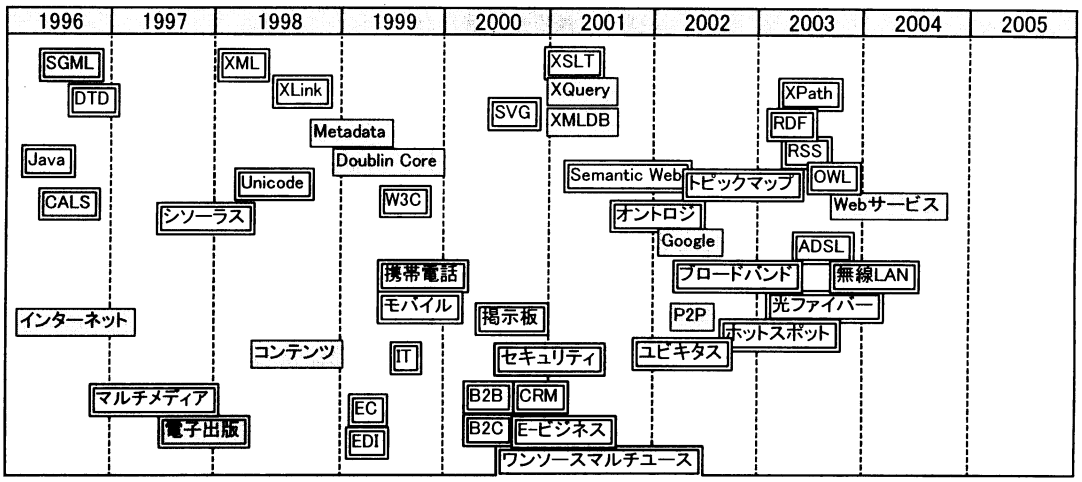

図 12　トピックワード年表

2001 年には，Semantic Web についての研究報告が行われ，その後のオントロジ，トピックマップ，OWL へと継続する流れとなる．

2002，2003 年にはインフラ系の用語が次々と登場する．これは日本でのブロードバンドの拡大を反映しており，研究報告では，情報インフラの整備に対して XML を用いた解決法などが提案されている．

6. まとめ

本研究では，10 年間のデジタルドキュメント研究の流れを俯瞰した．初期には文書の電子化や構造化が中心であったが，XML の登場と共にコンテンツに関する情報を対象とした，システムやサービスへと研究が変化した．

デジタルドキュメント研究において，XML は 1998 年の登場以降，中心と呼べる存在となり，情報記述手法として XML が一般的となった．最近の研究の対象は，XML で記述された情報に対する，EC や医療，教育などの各分野に特化した応用的処理手法に移り変わっている．

近年の特徴的な傾向として，携帯電話や Semantic Web に関する研究報告の増加がある．携帯電話は，端末機能・通信帯域共に今後も発展が見込まれ，デジタルドキュメントに対するインフラとして，大きな役割を占める可能性を秘めている．また，インターネットの普及に伴い，望んだ情報の検索・抽出が重要な課題となっており，その解決手段の一つとして Semantic Web やオントロジの研究が行われている．

今後のデジタルドキュメント研究では，XML に加え，携帯電話，Semantic Web が大きな役割を担っていくと予測する．

本研究で導き出したデジタルドキュメント研究 10 年間の傾向に対し，さらに外部の研究会などと比較することで，デジタルドキュメント研究の特徴を更に浮き彫りにし，10 年後のデジタルドキュメント研究の予測ができるのではないだろうか．

参考文献

[1] 斎藤伸雄，三田虎史，“デジタルドキュメント 10 年の傾向”，情報処理学会研究報告 Vol.2005，No.54，pp15-22，2005.

[2] “情報処理学会電子図書館”，情報処理学会，<http://www.ipsj.or.jp/05system/digital_library/index.html>，参照 2005-6-23

[3] “TRUE TELLER”,野村総合研究所，<http://www.trueteller.net/>，参照 2005-6-23

[4] “CiNii 論文情報ナビゲータ”,国立情報学研究所，<http://ci.nii.ac.jp/cinii/servlet/CiNiiTop#>，参照 2005-6-23.

[5] 藤井義信，清兼幸雄，中村正実，宮川純一，“ソフトウェア設計情報共有化の為のソフトウェア CALS システムに関する一考察”，情報処理学会研究報告 Vol.1996，No.76，pp9-16，1996.

[6] Tim Berners-Lee，“Semantic Web - XML2000”，W3C，<http://www.w3.org/2000/Talks/1206-xml2k-tbl>，参照 2005-6-23

[7] Frank Manola, Eric Miller eds,“RDF Primer”，W3C，< http://www.w3.org/TR/rdf-primer/>，参照 2005-6-23

[8] RSS-DEV Working Group，“RDF Site Summary (RSS) 1.0”，RSS-DEV WG，< http://web.resource.org/rss/1.0/spec>，参照 2005-6-23