

Journal of Japan Society of Information and Knowledge

情報知識学会誌

Vol. 20 No. 1 (Feb. 2010)

~~~~~目 次~~~~~

|      |                                                                                  |    |
|------|----------------------------------------------------------------------------------|----|
| 論 文  | Office Open XMLに対する長期署名の付与<br>山地一禎, 片岡俊幸, 宮地直人, 曾根原登 . . . . .                   | 1  |
| 論 文  | 異なる学問分野のコーパスを利用した専門用語抽出手法の提案<br>久保順子, 辻慶太, 杉本重雄 . . . . .                        | 15 |
| 調査報告 | 情報記録手法と記憶定着・理解度の関係についての実験報告<br>～手書き記録時とキーボード記録時の差異について～<br>曾根原士郎, 齋藤敦子 . . . . . | 32 |
| 解 説  | 著作物のフェアユースまたは公正使用に関する動向<br>時実象一 . . . . .                                        | 38 |
| 部会報告 | 平成21年度科学技術データベース懇談会報告<br>馬場哲也, 山下雄一郎, 岩田修一 . . . . .                             | 47 |

~~~~~お知らせ~~~~~

第7回 (2010) 論文賞投票公告
第18回情報知識学会年次大会論文募集
会員増加のための諸制度改定のお知らせ
投稿規定・執筆要項
事務局からのお知らせ

~~~~~



情報知識学会

<http://www.jsik.jp/>

# TOPPAN

ここには、人類の知と創造への  
エネルギーがあふれています。

そして、印刷とコミュニケーションの過去、現在、未来の姿を探る。

ここは人類の偉大なる知と創造へのエネルギーを感じることができるスペースです。



〒112-8531  
東京都文京区水道1丁目3番3号  
TEL: 03-5840-2300 (代)

<http://www.printing-museum.org/>

●交通:JRおよび地下鉄有楽町線、東西線、南北線、大江戸線飯田橋駅より徒歩約13分。地下鉄有楽町線江戸川橋駅より徒歩約8分。地下鉄丸の内線、南北線後楽園駅より徒歩約10分。●開館時間:10時～18時(入場は17時30分まで)●休館日:毎週月曜日(但し祝日の場合は翌日)、年末年始、展示替え期間●入館料:一般(中学生以上)300円、小学生100円、団体割引あり(税込)

論文

## Office Open XMLに対する長期署名の付与

### Application of the Long Term Signature to Office Open XML files

山地一禎<sup>1\*</sup>, 片岡俊幸<sup>1</sup>, 宮地直人<sup>2</sup>, 曾根原登<sup>3</sup>

Kazutsuna YAMAJI, Toshiyuki KATAOKA, Naoto MIYACHI, Noboru SONEHARA

1 国立情報学研究所 学術ネットワーク研究開発センター

R & D Center for Academic Networks, National Institute of Informatics

E-mail: yamaji@nii.ac.jp, kataoka@nii.ac.jp

2 有限会社 ラング・エッジ

LangEdge Inc.

E-mail: miyachi@langedge.jp

3 国立情報学研究所 情報社会相関研究系

Information and Society Research Division, National Institute of Informatics

E-mail: sonehara@nii.ac.jp

研究成果などの知的財産管理の手段として、電子署名とタイムスタンプ技術を基礎とした長期署名の応用が注目されている。本研究では、研究の過程で頻繁に利用される、オフィス系アプリケーションに対応した長期署名の付与方法を提案する。対象とするフォーマットは、マイクロソフト社のOffice 2007に採用されている、Office Open XML (OOXML) とした。長期署名フォーマットには、XMLで記述されるXAdESを採用した。OOXMLへのXAdESの実装として、標準長期署名と拡張長期署名と呼ぶ二種類の方法を提案した。前者はOOXMLで規定するXML署名のファイルに長期署名に関する記述を含める形式であり、後者は独立したファイルに記述する形式である。双方とも、OOXMLとXAdESの仕様を逸脱するものではない。この二種類の長期署名を、Office 2007が生成するOOXMLファイルに付与するアプリケーションを開発した。長期署名付きファイルをOffice 2007で開いた結果、拡張長期署名はXML署名としての検証には成功するのに対し、標準長期署名は失敗することが確認された。XML長期署名としては、標準長期署名の方がより一般的な実装方法であり、OOXMLにおいて長期署名を付与する書式を明らかにするとともに、Office 2007によるOOXMLの解釈にも対処が必要であることが示唆された。

The long term signature (LTS) is composed of an electronic signature and timestamp and it may be an effective technology for protecting intellectual property. This study describes a method of applying LTSs to files generated by office applications. The targeted files are in Office Open XML (OOXML) and are opened by MS Office 2007. XML LTS is described in XAdES format. Two types of LTS, that

is, standard LTS and expanded LTS, can be used to apply XAdES to OOXML. The standard format involves a LTS description in a XML signature file that is defined by OOXML. The expanded format separates the LTS description from the XML signature file. Both formats satisfy the OOXML and XAdES specifications. As a result of reading two XAdES signature word processing files by using Office 2007, the file with the expanded LTS was recognized as a valid file with an XML signature, whereas the file with the standard LTS was not.

キーワード: 長期署名, 電子署名, タイムスタンプ, 知的財産, 研究ノート

Long-term Electronic Signature, Electronic Signature, Timestamp, Intellectual Property, Lab Notebook.

## 1 はじめに

国立大学の法人化などを契機に, 大学における知的財産に対する認識が高まってきた[1]. その対象は, 研究成果の帰属や特許取得戦略といったポリシー的な側面のみならず[2], 研究ノートの導入によるリスクマネジメントといった実地的な内容まで多岐にわたる[3]. 一方で, インターネット上におけるデータベースや参加型サイトの普及に伴い, 従来は文献としてのみ公開されていた研究成果が, 現在では様々な段階・形式で公開されるようになってきた[4][5]. しかしながら, そうしたオープンサイエンス的な流れの中では, 知的財産保護に対するシステムのおよびポリシー的な解決策が提示されておらず, そうした問題を解決することは, 今後の普及に向けての重要な課題である.

一方, e-文書法の制定に伴い, 企業内では, 財務・税務関係の法定保存文書を電子的にセキュアに保存するための環境整備が進んでいる. このとき, 文書の作成者や作成時期が紙文書などと電子化した文書が同一であることが確認できること, すなわち, 文書の真正性を担保する

ために, 電子署名とタイムスタンプの付与が求められている[6]. ここで, 電子署名とは, 電子文書の作成者(署名者)が特定でき, その内容の非改ざん性を保障するための行為あるいはシステムのことである. これにより, 電子文書の「誰が」「何を」が保障される. さらに, 「いつ」電子文書が作成されたかを担保する規格がタイムスタンプである. こうした技術は, 事務的な文書保存以外にも, 先使用权を立証し知的財産を確保する目的でも有効利用されている[7]. 最近では, 文書の真正性をより長期的に担保する目的で, 電子署名とタイムスタンプ技術を組み合わせた長期署名が提案されており[8][9], JISにおいても標準化が行われた[10][11].

これまでに我々は, 学術分野においてアイデアや研究成果の先取権(先使用权)を主張する目的で発行されるプレプリントに対し, 長期署名を付与するシステムの開発を進めてきた[12]. プレプリントシステムで対象としたのはPDFファイルであったが, より汎用性を高めるためには, 通常の業務で頻繁に使われるオフィス系アプリケーションのファイルフォーマットにも対応することが望まれる. そうした長期署名技術の拡張は, 学術分野



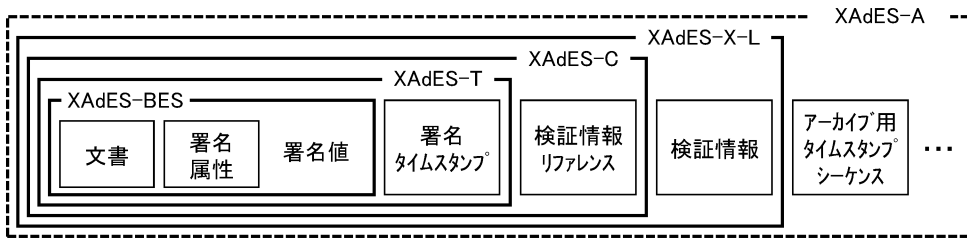


図 1 XAdES 長期署名フォーマット

における知的財産保護の条件を満たしつつ、研究者がインターネットを通じて研究成果を公開する自由度を確保することができる。現在、オフィス系アプリケーションとして多く利用されているものとしては、マイクロソフト社と OpenOffice.org が提供するものがある[13]。現在では、双方ともに ISO 化された XML 形式のファイルフォーマットと、XML 署名 (Xmldsig) [14] による電子署名を採用し、前者は Office Open XML (OOXML) [15]、後者は Open Document Format (ODF) [16] で記述される。ODF に関しては、Xmldsig を拡張した XML 長期署名 (XAdES) [9] の適用が試みられ、実際に、電子署名法に基づく認定認証業務の中で利用されている[17]。一方、よりシェアが高いと考えられる OOXML に関しては、その長期署名化に関するシステム実装の報告はこれまでなされていない。そこで本研究では、研究成果のセキュアな共有を可能にする基盤技術として、OOXML の長期署名化を実現することを目的とした。

## 2 長期署名

図 1 は、XAdES による長期署名フォー

マットを記述したものである[9]。ここで、XAdES-BES (XML Advanced Electronic Signatures - Basic Electronic Signature) の部分が電子署名に該当し、電子文書、署名属性、署名値から構成される。署名値は、電子文書から生成する固定長の疑似乱数であるハッシュ値を、署名者の秘密鍵で暗号化した値である[18]。署名文書の受取者は、このハッシュ値を利用して非改ざん性を検証する。ただし、署名を行った本人は内容を改変することが可能であることから、電子署名に加えて、時間軸での電子文書の存在を保障するタイムスタンプの付与が必要とされる[19]。図 1 の場合では XAdES-T (XAdES with Time) までのフォーマットが、これに該当する[9]。

電子署名やタイムスタンプは、いずれもデジタル署名技術を利用しており、アルゴリズムの脆弱化や鍵漏洩への対応として、有効期限や失効（無効化）機能を有している。すなわち、長期的に証明力を保持できない仕組みとなっており、研究成果のように長期にわたって存在証明と非改ざん証明の保証が必要とされる対象には適していない。これに対処するために提案されたのが長期署名であり、署名の検証に用いた証拠として、認証パ

ス上の全ての公開鍵証明書とその失効情報へのリファレンス（図 1 XAdES-C : XAdES with complete validation data references）と実際の検証情報（図 1 XAdES-X-L : Extended long electronic signatures with time）を付加し、長期署名フォーマット全体にアーカイブタイムスタンプ（図 1 XAdES-A : Archival electronic signatures）を付与する。このとき、XAdES-Cは、ETSI TS 101 903[9]およびJIS プロファイル[11]ともに、その付与を任意としている。本研究では、OOXMLで作成されるコンテンツに対し、XAdESによる長期署名を付与するための方法を提案する。

### 3 Office Open XML

OOXMLでは、画像などのバイナリデータを除けば他は全てXML記述されており、それらのファイルをZIP形式で圧縮して1つのファイルにまとめる。例えば、Office 2007 Wordで作成したdocx拡張子のZIPファイルを解凍すると、パッケージの内部は以下になっている。

- `_rels` (フォルダ)  
分散保存されたファイルの関連情報を保存
- `docProps` (フォルダ)  
プロパティに関する情報を保存
- `word` (フォルダ)  
文書本体およびスタイルに関する情報を保存
- `[Content_Types].xml` (XMLファイル)  
含まれる全てのファイルのデータタイプを定義

`word`フォルダは、エクセルの場合は`xls`、パワーポイントの場合は`ppt`フォルダと

なる。こうしたオフィスファイルへの電子署名には、`Xmldsig`を適用することができ、OOXMLの仕様書内にその方法が記述されている。`Xmldsig`を付与した場合には、上記のパッケージ構成に対して、

- `_xmsignatures` (フォルダ)

が追加され、必要な情報が保存される。例えば、Office 2007で最初に電子署名を付与した場合には、`sig1.xml`というファイルが生成され、`Xmldsig`の内容が保存される。なお、電子署名ファイルに適用されるスキーマは、OOXMLのものではなく、標準の`Xmldsig`となる。本研究では、この`_xmsignatures`に保存される署名情報を対象とし、XML署名(`Xmldsig`)からXML長期署名(XAdES)へと拡張する方法を提案する。

### 4 Office Open XMLへの長期署名

#### 4.1 OOXMLによるXmldsigとXAdESによる長期署名フォーマット

ETSI TS 101 903[9]によるXAdESフォーマットと各XML要素の説明を表 1に示す。○マークで示した要素はRFC3275による`Xmldsig`で必要とされており、●マークで示した要素は長期署名用に追加されたものである。XAdESによる長期署名の場合は、●1と●3の対応からわかるように、署名対象がObject全体(●2)にはならず、`SignedProperties`(●3)のみとなる。長期署名用のXAdES-A(●4)のように必要に応じて追加されていく内容は`UnsignedProperties`配下に記述して署名の対象外とすることで、XAdES-BES付与時のハッシュ値の永続性が担保できる構造となっている。

表 1 XAdES 署名の標準的な要素

XML 要素	説明
○ <Signature>	Xmldsig 開始タグ
○ <SignedInfo>	署名対象要素
○ <CanonicalizationMethod/>	署名対象正規化手法指定
○ <SignatureMethod/>	署名アルゴリズム指定
○ <Reference/>	署名対象への URI 指定 1 とオプション
○ :	変換手法指定署名アルゴリズム指定
○ :	署名対象は複数指定可能
●1 <Reference URI="# <u>xades</u> "/>	XAdES 署名対象への URI 指定
○ </SignedInfo>	署名対象要素終了
○ <SignatureValue/>	署名値 (Base64)
○ <KeyInfo/>	署名者の秘密鍵情報
○ <Object/>	署名対象内包時のオブジェクト要素
●2 <Object>	長期署名用オブジェクト (必須)
● <QualifyingProperties>	XAdES 要素開始
●3 <SignedProperties Id=" <u>xades</u> ">	XAdES 署名対象要素
● <SignedSignatureProperties/>	XAdES 署名要素 (例: <SigningTime>等)
● </SignedProperties>	XAdES 署名対象要素終了
● <UnsignedProperties>	XAdES 非署名対象要素
● <UnsignedSignatureProperties>	XAdES 非署名要素
● <SignatureTimeStamp/>	XAdES-T 署名タイムスタンプ要素
● <CertificateValues/>	XAdES-X-L 証明書一覧要素
● <RevocationValues/>	XAdES-X-L 検証情報一覧要素
●4 <ArchiveTimeStamp/>	XAdES-A 保管タイムスタンプ要素
● :	(複数回可能)
● </UnsignedSignatureProperties>	XAdES 非署名要素
● </UnsignedProperties>	XAdES 非署名対象要素終了
● </QualifyingProperties>	XAdES 要素終了
● </Object>	長期署名用オブジェクト終了タグ
○ </Signature>	Xmldsig 終了タグ

OOXMLによるXmldsigの例を表 2に示す。OOXMLでは、Xmldsigにて検証可能なフォーマットとして、△1で示したidPackageObjectで参照されるObject要素 (Package-Specific Object) を記述し、必要に応じて△2で示したidOfficeObjectで参照されるオプション Object 要素 (Application-Specific Object) を記述する。このとき、△1と△3および△2と△4の対応からわかるように、Reference要素で規定される署名の対象はObject全体になる。

#### 4.2 標準長期署名

OOXMLを長期署名化するためには、表 2のXmldsigに表 1の●で示したXAdESの要素を組み込めばよい。表 3に、その記述例を示す。◇1と◇4でPackage-Specific Objectを記述し、◇2と◇5でApplication-Specific Objectを記述することで、OOXMLにおけるXmldsigを構成している。一方、◇3と◇6では、XAdESによる長期署名を構成している。4.1で述べたように署名対象の指定は、Xmldsig

表 2 OOXML における署名例

	<?xml version="1.0" encoding="UTF-8" ?>
	<Signature Id="idPackageSignature" xmlns="http://www.w3.org/2000/09/xmldsig#">
	<SignedInfo>
	<CanonicalizationMethod Algorithm="http://www.w3.org/TR/2001/REC-xml-c14n-20010315" />
	<SignatureMethod Algorithm="http://www.w3.org/2000/09/xmldsig#rsa-sha1" />
△1	<Reference URI="#idPackageObject" Type="http://www.w3.org/2000/09/xmldsig#Object">
	<DigestMethod Algorithm="http://www.w3.org/2000/09/xmldsig#sha1" />
	<DigestValue>\$aDdLivsjfCQLGxtrqC0tWts0914=\$</DigestValue>
	</Reference>
△2	<Reference URI="#idOfficeObject" Type="http://www.w3.org/2000/09/xmldsig#Object">
	<DigestMethod Algorithm="http://www.w3.org/2000/09/xmldsig#sha1" />
	<DigestValue>60B0mxo/DLT80lprfcK9P/ieHGM=</DigestValue>
	</Reference>
	</SignedInfo>
	<SignatureValue>gZDjjpwCkFGZUz9X/O7cq\$6E54flNehWqz...</SignatureValue>
	<KeyInfo>
	:
	</KeyInfo>
△3	<Object Id="idPackageObject" xmlns:mdssi="http://schemas.openxmlformats.org/package/2006/digital-signature">
	:
	</Object>
△4	<Object Id="idOfficeObject">
	<SignatureProperties>
	<SignatureProperty Id="idOfficeV1Details" Target="#idPackageSignature">
	<SignatureInfoV1 xmlns="http://schemas.microsoft.com/office/2006/digsig">
	:

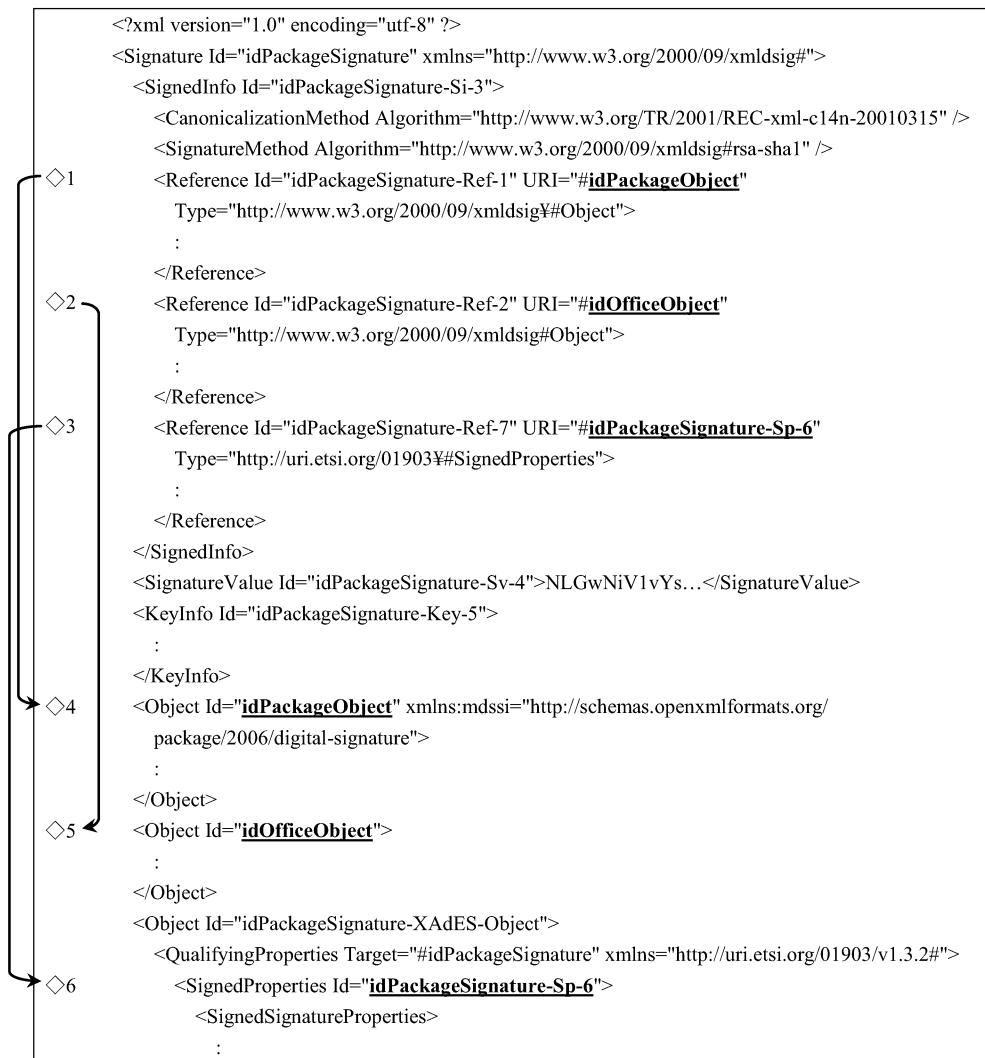
のようにObject全体ではなく、その子要素のSignedPropertiesに特定している。以上のように、一般的な方法として、XmldsigとXAdESの記述を単一のファイルで記述する方法を標準長期署名と呼ぶ。

#### 4.3 拡張長期署名

Office 2007ではXmldsigを付与することができるため、この記述とはできる限り干渉しないように、長期署名を付与する方法が考えられる。XAdESによる電子署名部分、すなわちXAdES-BESでは、表1の●3からはじまるSignedPropertiesの内

容もハッシュ値の計算対象となる。したがって、本文が同じ場合でもXmldsigとXAdES-BESでは署名値が異なる。そこで、Office 2007によるXmldsigはそのまま保存し、XAdESに関わる内容を独自のファイルに保存する実装方式を提案する。表4に、XAdES用ファイルの記述例を示す。□1で示したReferenceでは、Xmldsigのみを記述した署名ファイルをURIとして指定する。□2で示したReferenceでは、長期署名のためのObjectのSignedProperties (□6)を参照する。SignedInfo以降には、SignatureValue (□3)、KeyInfo (□4)

表 3 標準長期署名の署名例



を記述する。SignatureValueの署名値はXmldsigとして計算したものとは異なり、表 4のファイルが規定するXAdES-BESとして、長期署名のSignedPropertiesを署名対象に加えて計算した署名値とする。KeyInfoは、Xmldsigと同じものを利用する。XAdESのObject (◇5) には、表 1の●2以降のObjectの内容を記述する。以上に従い、Xmldsigとは分離したファイル

でXAdESによる長期署名を行う方法を拡張長期署名と呼ぶ。

#### 4.4 アプリケーション構築

OOXMLの長期署名フォーマットとして本研究で提案する標準長期署名と拡張長期署名を実際に付与し、本研究で提案している仕様の妥当性を検討するためのアプリケーションを開発した。長期署名

表 4 拡張長期署名の長期署名部記述例

	<?xml version="1.0" encoding="UTF-8" ?>
	<Signature Id="idXadesSignature" xmlns="http://www.w3.org/2000/09/xmldsig#">
	<SignedInfo Id="idXadesSignature-Si-2">
	<CanonicalizationMethod Algorithm="http://www.w3.org/TR/2001/REC-xml-c14n-20010315" />
	<SignatureMethod Algorithm="http://www.w3.org/2000/09/xmldsig#rsa-sha1" />
□ 1	<Reference Id="idXadesSignature-Ref-1" URI="_xmldsignatures/sig1.xml">
	<Transforms>
	<Transform Algorithm="http://www.w3.org/TR/2001/REC-xml-c14n-20010315" />
	</Transforms>
	<DigestMethod Algorithm="http://www.w3.org/2000/09/xmldsig#sha1" />
	<DigestValue>IAGQ8Mz2NchhdS/cCPiNsFoHw7w=</DigestValue>
	</Reference>
□ 2	<Reference Id="idXadesSignature-Ref-6" URI="#idXadesSignature-Sp-5"
	Type="http://uri.etsi.org/01903#SignedProperties">
	<Transforms>
	<Transform Algorithm="http://www.w3.org/TR/2001/REC-xml-c14n-20010315" />
	</Transforms>
	<DigestMethod Algorithm="http://www.w3.org/2000/09/xmldsig#sha1" />
	<DigestValue>wIPeYDJJXmFolW8N+ohgaUQtE+8=</DigestValue>
	</Reference>
	</SignedInfo>
□ 3	<SignatureValue Id="idXadesSignature-Sv-3">enf7MqAWThKHpLbrue+cMIL&uP...
	</SignatureValue>
□ 4	<KeyInfo Id="idXadesSignature-Key-4">
	:
	</KeyInfo>
□ 5	<Object Id="idXadesSignature-XAdES-Object">
	<QualifyingProperties Target="#idXadesSignature" xmlns="http://uri.etsi.org/01903/v1.3.2#">
□ 6	<SignedProperties Id="idXadesSignature-Sp-5">
	:

は、XAdES-Cを含めない、XAdES-BES、XAdES-T、XAdES-X-L、XAdES-Aで構成した。任意の電子証明書およびタイムスタンプ局を利用して、XAdESによる長期署名が付与できるシステムとするために、以下の機能を実装した。

- Windows証明書ストアからの証明書選択機能
- タイムスタンプ局指定機能
- 長期署名付与機能  
(標準長期署名、拡張長期署名を選択可能)

- 保管タイムスタンプの追加機能
- 長期署名検証機能  
(Xmldsig, XAdESのスキーマ検証を含む)

これらの機能を、Microsoft Visual C++を利用して、Xmldsig付与に関する研究開発でもしばしば利用されている.NET Framework[20][21]上で動作するアプリケーションとして開発した。XAdESに関する基本的な処理は、Le-XAdESを利用した[22]。ZIP圧縮の処理には、zlib[23]とInfo-ZIP[24]を利用した。開発したアプリ



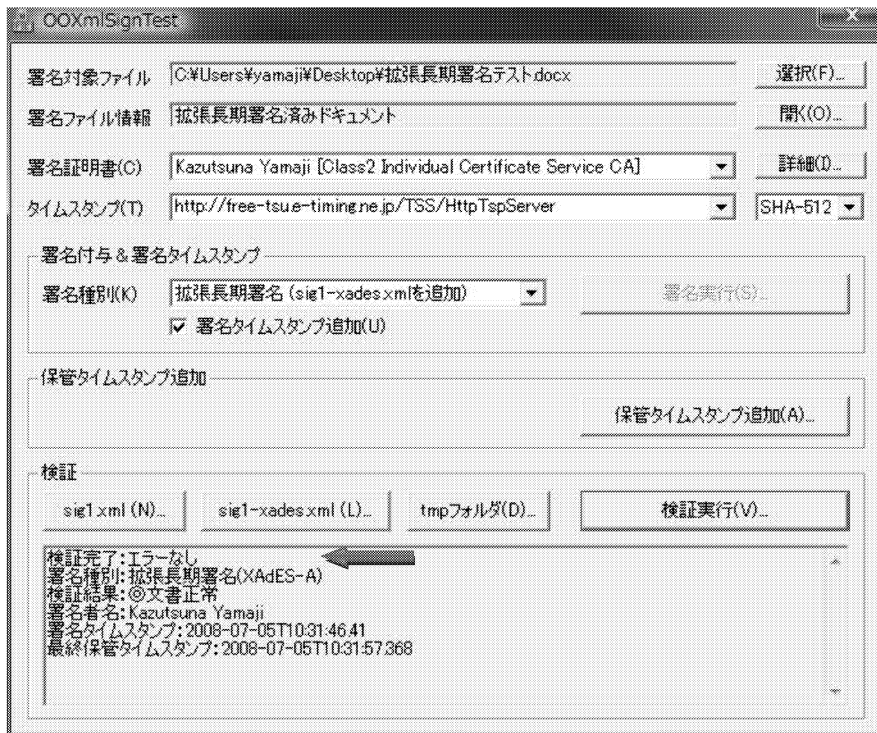


図 2 XML 長期署名付与アプリケーションのスクリーンショット

ケーションのスクリーンショットを図 2 に示す. 長期署名では, 図 1 における XAdES-BES, XAdES-T, XAdES-A が検証の対象となるが, 本アプリケーションにおいても, 以下のように各内容が個別に検証されるように構築した. XAdES-A に関しては, 全ての世代のタイムスタンプが検証されるようにした.

- Xmldsig の検証 (XAdES-BES)
- 署名タイムスタンプの検証 (XAdES-T)
- 全世代のアーカイブタイムスタンプの検証 (XAdES-A)

## 5 評価と考察

OOXML は標準化された仕様であるが,

現在のところ, 技術開発のために汎用的に利用されている検証ツールは存在しない. そこで本研究では, Office 2007 Word を利用し, ベースとなる文書ファイルの作成, および, 電子署名に関する動作確認を行った. まず, Office 2007 Word で文書ファイルを作成した後に, 開発したアプリケーションで標準長期署名および拡張長期署名を付与し, どちらも Xmldsig および XAdES として正常に検証できることを確認した (図 2 中赤矢印). 電子証明書<sup>1</sup>と時刻認証局<sup>2</sup>には, 第三者機関が提供するサービスを利用した.

<sup>1</sup> 日本ベリサイン株式会社 Class2 スタンドダード

<sup>2</sup> アマノタイムビジネス株式会社 AMANO Time Stamp Service Type-Free-A

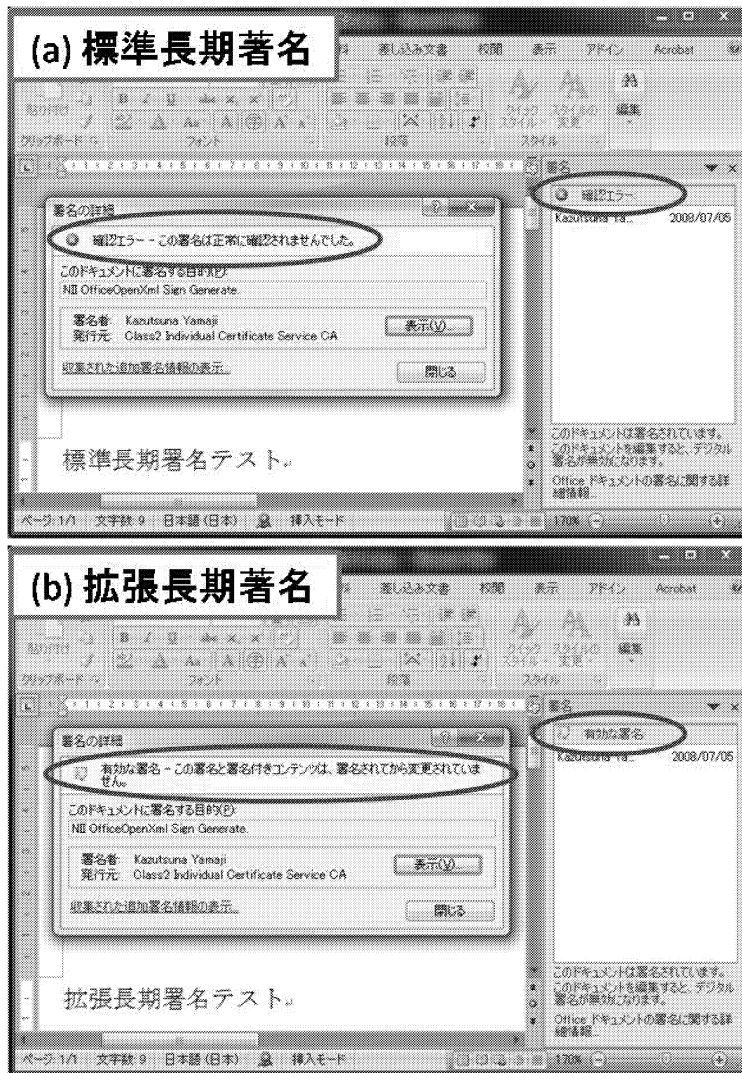


図 3 Office 2007 Word ファイルに対する標準および拡張長期署名付与結果

各長期署名を付与したファイルをOffice 2007 Wordで開いた結果を図3に示す。これより、標準長期署名は、本研究で開発したアプリケーションでの検証には成功するが、Office 2007の応答としてはエラーとなる。これに対し、拡張長期署名は、Office 2007においても有効なXmlsigといして認識されていることがわかる。また、OOXMLは複数のXMLファイルをZIP

形式で圧縮した構造となっているが、Office 2007では構造の詳細なチェックはせずに、拡張長期署名でXAdESを追加したファイルの添付も許容されている。このとき、検証の対象となるのは、XAdESを含まないXmlsigの部分である。すなわち、Office 2007においても、拡張長期署名では、署名値は正しく本文は改ざんされていないことが確認できる。

標準長期署名におけるOffice 2007の応答のエラーの原因に関しては、以下の理由が考えられる。まず、OOXMLの電子署名に適用されるスキーマはXmldsigであり、本研究で開発したアプリケーションにおいて電子署名値とスキーマの検証には成功しており、有効なXmldsigが付与されていると評価できる。従って、Office 2007においてもXmldsigとしての検証に成功することが期待されるが、実際にはエラーとなる。OOXMLの仕様書では、SignedInfo 要素には、Package-Specific ObjectとApplication-Specific Object要素へのReferenceが指定可能であるが、それ以外のReferenceの指定に関する制約は書かれていない。しかしながら、図3の結果より、Office 2007の動作としては、Package-Specific Object と Application-Specific Object要素以外のReferenceは受け付けないと推定される。これはXmldsigの仕様やスキーマにも記述されていない制限である。こうした動作が、Office 2007の実装上の不備に起因するものなのか、仕様上のものなのかは、現在のソフトウェア流通メカニズムでは判別することが困難である。ただし、ここで得られた結果を基に、標準長期署名のような形式でOOXMLに対して長期署名を付与するためには、Object要素だけではなく、その下位に位置するSignedProperties要素(表3◇3)に対しても、Referenceを許可するようにOOXMLの仕様を厳密化する必要がある。

本研究では、OOXMLの仕様から派生する長期署名方法として標準長期署名と拡張長期署名を提案した。これらの方法とは別に、OOXMLの仕様とは全く別に

長期署名を付与することも可能である。すなわち、標準長期署名により作成されるファイルを、OOXMLが解釈できない場所あるいはファイル名で保存すればよい。この長期署名の付与方法を独立長期署名と呼ぶことにする。標準長期署名、拡張長期署名および独立長期署名は、それぞれXAdESにより定義される長期署名フォーマットから逸脱するものではない。以下では、各方式の得失について考察する。

拡張長期署名は、表4中□1で示したようにXmldsigに対するリンクをもつことになる。したがって、Xmldsigに利用された電子署名が長期署名に利用されたものと等しい場合には、表4のXAdESファイルの中でその検証情報(図1のXAdES-CとXAdES-X-T)を保存することができる。しかしながら、双方の電子署名が異なる場合には、その検証情報を保持することができないという制約がある。

OOXMLに従う仕様としては、署名情報を1つのファイルに保存する標準長期署名がXAdESの表現として最も標準的である。拡張長期署名は、XmldsigとXAdESを共存させるという点で冗長性があるが、現状のOOXMLの仕様の拡張版と捉えることができる。一方、独立長期署名は、OOXMLの仕様とは関連性がない。

Office 2007がXmldsigとして正常に認識できるのは、拡張長期署名だけである。標準長期署名は、図3(a)に示したように、Xmldsigとしてはエラーとなる。独立長期署名は、OOXMLの仕様から逸脱することを目的とした方法であるため、Office 2007のXmldsigとしては、評価の対象外と

なる。独立長期署名では、Office 2007の既存の機能としては、いかなる署名も付与されていないと認識されることになる。

以上の比較により、Office 2007の改良が可能であれば、OOXMLの長期署名方法としては標準長期署名が最も優れた方法であると考えられる。ただし、既存のOffice 2007ユーザの利便性を考慮すると、現状では、拡張長期署名の利用が優位なケースも考えられる。

## 6 おわりに

本研究では、オフィス系アプリケーションのためのISO標準フォーマットである、OOXMLへの長期署名方式について検討した。OOXMLから派生する方法として、標準長期署名と拡張長期署名を提案し、それぞれの特徴について評価した。現状のOOXMLの仕様は、Xmldsigを付与することにのみ対応しているが、本研究の成果は、OOXML においてXAdESを実装する上での仕様拡張の方向性を与えるものである。また、OOXMLの仕様を逸脱しないと考えられる標準長期署名が、Office 2007ではXmldsigとしてエラーになることから、OOXMLの厳密化の必要性、あるいは、Office 2007の改善点が示唆された。

インターネット上に流通する研究成果としては、これまではPDFフォーマットで公開される文献が中心であった。今後は、そのモデルを軸として、文献に関連する補遺的なコンテンツを、多様なフォーマットで配信するためのインターネット上のサービスやコミュニティの育成が進んでくると予想される[25]。本研究の

成果は、そうした状況においても、コンテンツの流通を妨げることなく知的財産保護を確保するための、基盤技術を提供するものである。今後は、長期署名付きコンテンツの流通と、その情報を利用したサービスについて検討を進め、よりリッチな学術成果共有社会の育成に寄与したいと考えている。また、本研究の一部をワーキンググループにて報告した次世代電子商取引推進協議会では、その内容が高く評価されており[26]、OOXMLに対する長期署名方法の標準化という側面からも、今後、産学連携により本研究の成果を展開していく予定である。

## 参考文献

- [1] Shimada, R.: “Intellectual Property Management of National University Corporations Shift to Institutional Ownership and its New Challenges” Vol.1, No.1, 37-45, 2005.
- [2] 青山紘一: 「研究成果の帰属と補償 (下) —大学・公的研究機関における研究成果の帰属と管理—」, 情報管理, Vol.45, No.12, 845-857, 2003.
- [3] 妹尾堅一郎: 「研究ノートのリスクマネジメント」, 研究・技術計画学会, Vol.21, No.2, 435-438, 2006.
- [4] 西村美咲, 横井茂樹, 安田孝美: 「研究室の情報共有・公開を支援するCMSを基盤としたWebシステムの構築」, 情報処理学会研究報告, Vol. IS-103, No.4, 23-28, 2008.
- [5] Yamaji, K., Sakai, H., Okumura, Y. and Usui, S.: “Customizable neuroinformatics database system: XooNlps and its

- application to the pupil platform”,  
Computers in biology and medicine,  
Vol.37, No.7, 1036-1041, 2006.
- [6] 石島正勝: 「情報の権限付与と共有範囲を考える」, 情報の科学と技術,  
Vol.56, No.1, 19-24, 2006.
- [7] 池畠裕介: 「先使用権制度の円滑な利用に関する調査研究」, 知財研紀要,  
Vol.16, No.11, 1-8, 2007.
- [8] Pinkas, D., Ross, J. and Pope, N.: “Electronic Signature Formats for Long Term Electronic Signatures”, IETF RFC3126, 2001.
- [9] European Telecommunications Standards Institute: “XML Advanced Electronic Signatures (XAdES)”, ETSI TS 101 903 v1.3.2, ETSI, 2006.
- [10] 日本工業規格: “CMS利用電子署名 (CAdES)の長期署名プロファイル”, JISX5092, 2008.
- [11] 日本工業規格: “XML署名利用電子署名(XAdES)の長期署名プロファイル”, JISX5093, 2008.
- [12] 山地一禎, 片岡俊幸, 曾根原登, 行木孝夫: 「プレプリントへの長期署名付与および検証システムの構築」, 情報知識学会誌, Vol.18, No.3, 240-248, 2008.
- [13] Kosek, J.: “From the Office Document Format Battlefield”, IT Professional, Vol.10, No.3, 51-55, 2008.
- [14] Eastlake, D., Reagle, J. and Solo, D.: “XML-Signature Syntax and Processing”, IETF RFC3275, 2002.
- [15] International Organization for Standardization: “Information technology - Document description and processing languages - Office Open XML File Formats”, ISO/IEC 29500, 2008.
- [16] International Organization for Standardization: “Document Format for Office Applications (OpenDocument)”, v1.0, ISO/IEC 26300, 2006.
- [17] 本間史夫: 「証明書電子更新サービス」, PKI-Jジャーナル, Vol. 2007最終号, 103-104, 2007.
- [18] Housley, R.: “Cryptographic Message Syntax”, IETF RFC3852, 2004.
- [19] Adams, C., Cain, P., Pinkas, D. and Zuccherato, R.: “Internet X.509 Public Key Infrastructure Time-Stamp Protocol (TSP)”, IETF RFC3161, 2001.
- [20] Bull L., and Squire D. M.: “XML Signature Extensibility Using Custom Transforms”, Lecture Notes in Computer Science, Vol.3306, pp.102-112, 2004.
- [21] Mahmud M.: “A Study on Data Security in the Microsoft .NET Platform”, Proc. National Conference on Cryptography and Information Security, NCCIS 2006, pp.45-49, 2006.
- [22] 宮地直人: “Le-XAdESライブラリ, 有限会社ラング・エッジ”, 入手先 <http://www.langedge.jp/pub/LeXAdES/>, 参照2009-04-07.
- [23] Roelofs, G., Gailly, J. and Adler, M.: “zlib software”, available from <http://zlib.net>, accessed 2009-04-07
- [24] Adler, M., Bush, J., Davis, K., Denker, H., Dubois, J., Gailly, J., Goatley, H., Gordon, E., Gorman, I., Herborth, C., Haase, D., Hartwig, G.,

- Heath, R., Hudson, J., Kienitz, P., Kirschbaum, D., Lee, J., Linden, O., Mandrichenko, I., Miller, S., Monesi, S., Owens, K., Petrov, G., Roelofs, G., Rommel, K., Salisbury, S., Smith D., Schweda, S., Spieler, C., Truta, C., Verheijen, A., Behren, P., Wales R. and White, M.: “Info-ZIP”, available from <http://www.info-zip.org>, accessed 2009-04-07.
- [25] Brunger-Weilandt, S.: “E-Science - Advancing new ways of scientific communication”, Information Services and Use, Vol.27, No.4, 161-166, 2007.
- [26] 次世代電子商取引推進協議会: “電子署名普及に関する活動報告2008”, 入手先 <http://www.ecom.jp/results/h20seika/H20results-14.pdf>, 参照2009-10-06.  
(2009年4月9日受付)  
(2009年12月18日採択)  
(2009年12月22日オンライン公開)

論文

# 異なる学問分野のコーパスを利用した専門用語抽出手法の提案 Automatic Term Recognition Using the Corpora of the Different Academic Areas

久保順子<sup>1\*</sup>, 辻慶太<sup>1</sup>, 杉本重雄<sup>1</sup>

Junko KUBO\*, Keita TSUJI, Shigeo SUGIMOTO

1 筑波大学大学院 図書館情報メディア研究科

Graduate School of Library, Information and Media Studies, University of Tsukuba

〒305-8550 茨城県つくば市春日1-2

E-mail: {kuboj35, keita, sugimoto}@slis.tsukuba.ac.jp

コンピュータを使用した専門用語自動抽出は、従来、対象とする専門分野のテキストコーパスのみをデータとして行っているものが多かった。しかし、専門用語の特徴として、対象分野のコーパスに頻出し、対象分野以外の他分野コーパスにはあまり多く出現しない点が挙げられる。そこで本研究では、対象分野コーパスと他分野コーパスとの用語の出現率の差を考慮した手法を提案する。実験では、女性学のテキストを対象分野のコーパスとして使用し、他分野のコーパスとして 39 分野のテキストを使用した。実験の結果、従来の代表的手法よりもかなり高い精度で用語が抽出できることが明らかとなった。また 39 分野のテキストから任意のテキストを選び他分野コーパスとして用いてコーパスの規模を縮小できるか実験を行った。その結果、対象分野と類似した分野のテキストを用いることで、39 分野すべてのテキストを用いた場合の抽出精度・再現率に近づくことができた。

In this paper, we propose a method for automatic term recognition (ATR) which is using the statistical differences of relative frequencies of terms in target domain corpus and in others. The target terms more frequently appear in target domain corpus than in other domain corpus. Utilizing such characteristics will lead to the improvement of extraction performance. Most of the ATR methods proposed so far only use the target domain corpus and do not take such characteristics into account. For the extraction experiment, we used the abstracts of the *Women's Studies International Forum* as a target domain corpus and those of academic journals of 39 domains as non-target domain corpus. The extraction performance was examined and we found that our method outperformed the existing ATR methods. We confirmed that it is possible to decrease the size of the other domain corpus by the experiments which used random journals out of 39 domains. As a result, we found that we used some corpus consists of journals which is similar to target domain is almost as high extraction performance as the corpus consists of 39 journals.

キーワード: 専門用語, 専門用語自動抽出, 頻度, コーパス, 女性学

Terminology, Automatic Term Recognition, Term Frequencies, Corpus, Women's Studies

## 1 はじめに

従来、専門用語辞書の作成と更新には、専門家による人手や時間など多大なコストが必要であった。コンピュータを使用した専門用語の自動抽出 (Automatic Term Recognition) は、専門用語辞書の迅速な作成・更新に役立つ技術であり、これまでに様々な手法が提案されてきた。従来のテキストコーパスからの専門用語自動抽出では、対象とする専門分野のテキストコーパスのみを情報源として用い、そこでの候補用語の出現頻度・出現パターンなどに基づいて抽出が行われ、対象とする専門分野以外の学問分野 (以下、他分野と表記) のテキストは情報源として用いられないことが多かった。しかし、Uchimoto et al.が指摘するように、専門用語には、(1) 対象とする専門分野 (以下、対象分野) の文書に頻出する、(2) 対象分野の common words ではないことが多い、(3) 対象分野以外のコーパスにはあまり多く出現しない、という特徴がある[1]。従って、従来手法のように対象分野のコーパスのみを利用するのではなく、対象分野と他分野それぞれにおける候補用語の出現率を比較し、対象分野における出現率が相対的に高い語を抽出するのが有効であると思われる。本研究ではそのように異なる専門分野のコーパスそれぞれにおける出現率の差に基づいて専門用語を抽出する手法を提案する。

対象分野のコーパスと他のコーパスを用いて、用語の出現頻度の差などを比較し専門用語抽出を行っている研究は他にもある。本研究では、次章に述べるこれら関連研究との比較により主に以下の3点に留意した。すなわち、(a) 他分野のコーパスについても、一般的なテキストではなく、学術的なテキストとすること、(b) 既存の代表的な専門用語自動

抽出手法との比較評価を行うこと、(c) 統計学的な裏付けを持ち、意味が理解しやすい尺度を提案すること、である。

(a) について述べると、学問分野のテキストと、一般的なテキストとにおける専門用語の出現頻度を比較した場合[2]、学術論文に頻出する用語 (例えば, “study”, “research” など) が専門用語として抽出されてしまう問題がある。そのため本研究では、専門分野以外の学問分野のテキストを、他分野のコーパスとして用いた。(b) に関しては、関連研究[1, 2, 3, 4, 5]では行われていないが、本研究では従来の専門用語抽出手法と比較することで、他分野のコーパスを用いる有効性を示す。(c) に関しては、本研究の尺度は後述のように統計学的基盤を持つものであり、意味の解釈が容易であるという長所を挙げることができる。

本稿では、上記の(a), (b), (c)に焦点を当て、専門用語抽出の精度・再現率を算出する実験を行い、既存の専門用語抽出手法と比較することで、提案手法の有効性を示す。さらに、対象分野と類似した分野のテキストを他分野コーパスとして用いることによって、他分野コーパスの規模を縮小することができるかについても検討する。

以下、専門用語抽出の関連研究については2章で、本実験に用いたデータに関しては3章で、提案手法および既存手法の説明は4章で、実験結果については5章で、考察は6章で述べ、7章で総括する。

## 2 関連研究

これまでの専門用語抽出研究に関するレビューは、Kageura & Umino[6]に詳しいが、専門用語自動抽出手法に関連が深い研究として、自動索引法 (Automatic Indexing) がある。自



動索引法は、電子化されたテキストに対して、自動的に索引語を付与する手法である。それらを含めると、本研究に関連する手法としては、(1) 品詞パターンに基づく手法、(2) 語構成要素の接続に注目した手法、(3) 語長と（単一分野のコーパスにおける）頻度に基づく手法、(4) 複数分野のコーパスにおける頻度に基づく手法、の4つがある。以下では、それぞれについて説明する。

(1) 品詞パターンに基づく手法：候補用語の品詞や形態素など言語学的な観点で抽出精度の向上をめざした研究[7, 8, 9]、特に語根や接辞に着目した形態論を展開した研究[10]、フレーズを構成する語数とその出現頻度とを考慮に入れた研究[11, 12]、N-gram の計算を用いた研究[13]、および用語変化を専門用語抽出手法に組みこんだ研究がある[14]。用語変化とは、(a) 正字法に基づくもの（ハイフンやスラッシュ、大文字小文字などの違い）、(b) 用語の形態に基づくもの（単数形・複数形など）、(c) 語彙に基づくもの（同義語など）、(d) 構造に基づくもの（前置詞の位置や変化：たとえば、clones of human と human clones など）、(e) 頭字語や略語、である。これらは、品詞パターンに基づく単純な手法であるため、単独で本研究との比較対象にはしない。ただし本研究における候補用語抽出の段階で、これらの研究で示された知見を取り込み、抽出する品詞パターンを構築し採用した。

(2) 語構成要素の接続に注目した手法：対象とする専門分野のコーパスを情報源とし、用語候補に隣接した語構成要素の頻度を考慮する手法である。これは多くの複合名詞を構成する用語は専門用語であるという仮説・前提に基づいている。この提案手法の代表例としては中川らの *FLR*[15]に基づく手法がある。

*FLR* は多くの文献で引用され、比較実験に用いられている。

(3) 語長と（単一分野のコーパスにおける）頻度に基づく手法：単一分野のコーパス中に出現した用語の頻度をベースにした代表的な手法として Frantzi et al. の *C-value*[16]が挙げられる。*C-value* は、1語で構成される用語については、専門用語候補としない。*C-value* もまた非常に多くの文献で引用され、比較実験に用いられている。

(4) 複数分野のコーパスにおける頻度に基づく手法：本研究と同様に、統計学的な裏付けを持つ代表的な研究として、確率分布のうちカイ二乗分布およびポワソン分布を応用した手法がある。長尾らは、ある用語の出現確率が全分野コーパスを通じて等しいかどうかを、カイ二乗値の算出によって判断し、その中から頻度に偏りがあるものを専門用語として抽出する手法を提案している[3]。このカイ二乗の値は、対象分野のみに偏って出現する語だけが低い値を持つのではなく、（対象分野の他に）他のいくつかの分野でも偏って多く出現する語も低い値を持ってしまう。従ってある1つの分野に関する専門用語を抽出するのには不向きである。また、Harter が提案する 2-Poisson Model [4, 5]は、対象分野と比較分野という2つの分野のコーパスを用い、それら2分野のテキストにおける対象用語の出現確率を推定しようとするものである。

Harter の手法は比較分野が1つに限定される。それに対して本研究手法は2つ以上の比較分野を統一的に扱うことができる利点がある。

この他に、複数分野のコーパスにおける頻度を用いるヒューリスティックな手法についても提案されている。Anselmo et al. は、2種類のコーパス中での候補用語の出現頻度に焦点を当てたヒューリスティックな尺度を提案

している。候補用語の出現頻度に加えて、比較コーパス中でも高頻度であるものや、対象コーパス中で DF（文書頻度）の割合が非常に少ない場合は除いている[2]。また、Uchimoto et al.は、対象分野コーパスとして、人工知能分野のデータコレクションを使用し、他分野コーパスとして毎日新聞のデータベース、NACSIS Academic Conference Database を使用している。Uchimoto et al.は、実験を通して、適合率と再現率の調和平均である F 値の最も高かったパラメータで、候補用語の出現頻度や出現分野に基づいた抽出尺度を決定している[1]。後者は、ヒューリスティックな手法であるが、比較コーパスが本研究と同様に複数の学問分野で構成されているため、比較手法としてふさわしいと思われる。そのため、本稿では、以上に挙げた(2) 中川らの FLR、(3) Frantzi et al.の *C-value* と、(4) Uchimoto et al.の UC[17]を実験の比較対象とする。

### 3 データ

本研究では、女性学分野を対象として専門用語抽出実験を行った。女性学分野を選んだ理由は 2 つある。1 つは、第一著者が女性学に関する知識を有しており、専門用語抽出における正解集合の作成を正確・容易に行えるためである。実際に女性センターのライブラリアン等と共に、女性情報研究会の一員として共同研究を進めており、女性学関連のデータを扱っている。本研究において、コーパスから人手によって正解用語を同定する作業が含まれるが、これについては第一著者が中心に行い、万全を期すために同研究会員らによるチェックも行った。

もう 1 つは、専門用語自動抽出を足がかりとした用語辞書作成のニーズが高いためで

ある。女性学分野は、(1) 学問分野として比較的歴史が浅く（アメリカにおける女性学の誕生は 1960 年代後半であり、女性学の専攻プログラムを持つ大学が増加したのは 1970 年代以降である）、既存の専門用語辞書があまり多くない、(2) 他の既存の学問分野に対して比較的小規模な学問分野であることから、辞書の刊行に対してコストをかけるだけの商業的な成功は見えにくく、また編著者のマンパワーも限られている（実際、本研究が用いた後述の辞書は、初版から改訂版までおよそ 10 年を要している）。よって、専門用語自動抽出が有効な分野の例として最適と考え、女性学分野を選んだ。

#### 3.1 コーパス

本研究では、2 つのコーパスを用いる。1 つはそこから専門用語を抽出する対象分野コーパスであり、もう 1 つは専門分野の用語以外の用語を多く含む他分野のコーパスである。

##### 3.1.1 女性学分野のコーパス

本研究の専門用語抽出対象コーパスは、女性学に関するジャーナル“*Women's Studies International Forum*”の抄録テキスト（以下、WSIF コーパス）である。

本研究では、1982年～2007年の1,212件の抄録をWSIFコーパスとして用いた[18]。含まれていた語は、延べ約18万語であった。

##### 3.1.2 他分野のコーパス

1981年～2007年の間に1,212件以上の電子化された抄録が入手可能な39ジャーナルを対象とした[19]。使用したジャーナル名は付録に挙げた通りである。各ジャーナルから1,212件の抄録を無作為抽出し、WSIFコーパスと同じ抄録数にすることでコーパスサイズを合わせ、論文の刊行年が偏らないようにし

た．全てのジャーナルを合わせて，含まれていた語は，延べ約 700 万語であった．以下ではこれを他分野コーパスと表記する．

### 3.2 候補用語

上記 2 種類のコーパスに対して，Brill's tagger[20]でタグ付けを行い，以下の品詞パターンを持っている最長単語列を候補用語とした．

{(形容詞|名詞)\*名詞+} または  
{(形容詞)\*(名詞)+(前置詞)\*(名詞)+}

ここで：

\*：0 回以上の繰り返し

+：1 回以上の繰り返し

である．

上記のパターンに該当する WSIF コーパス中の異なり用語は 20,266 語であり，他分野コーパス中の異なり用語は 630,541 語であった．

### 3.3 正解用語集合

本研究における正解用語集合とは，専門用語抽出の対象とした女性学に関する用語の集合である．正解集合として，WSIF コーパスから人手で抽出した語，および女性学専門用語辞書の見出し語の 2 つを用いた[21]．以下でそれぞれについて述べる．

#### 3.3.1 WSIF コーパスから人手で抽出した語

WSIF コーパスから，女性学の専門用語と思われるものを第 1 著者が手作業で抽出し，2 名の専門家によるチェックを受け，「正解集合 M」を作成した．用語総数は，2,104 語であった．

#### 3.3.2 女性学専門用語辞書の見出し語

女性学分野の辞書である“Women's Studies Encyclopedia” 1989-91 年版[22]および 1999 年

版[23]の Index に掲載されている 3,312 語[24]から WSIF コーパスに現れない語 (2,188 語)を除く「正解集合 D」を作成した．用語総数は 1,124 語であった．

## 4. 専門用語抽出手法

### 4.1 提案手法 MDP

MDP (*Minimum of the Difference between Population Proportions*) は，下記の(2)より，母比率の差の検定[25]の計算式をベースにしておき，統計的な基盤を持つものである．MDP は以下のように定義する．

$$MDP(T) = \min Z_i \quad 1 \leq i \leq N \quad (1)$$

$$Z_i = \frac{\frac{f_0(T)}{W_0} - \frac{f_i(T)}{W_i}}{\sqrt{\pi_i(T)(1-\pi_i(T))\left(\frac{1}{W_0} + \frac{1}{W_i}\right)}} \quad (2)$$

ここで：

$T$ ：候補用語

$N$ ：他分野コーパスを構成する全ジャーナルの数，本研究では 39

$W_0$ ：WSIF コーパス（対象分野）の候補用語の総延べ語数

$W_i$ ： $i$  番目の他分野コーパスの候補用語の総延べ語数

$f_0(T)$ ：候補用語  $T$  の WSIF コーパスでの出現頻度

$f_i(T)$ ： $i$  番目の他分野コーパスでの  $T$  の出現頻度 ( $1 \leq i \leq 39$ )

である．

また， $\pi_i(T) = \frac{f_0(T) + f_i(T)}{W_0 + W_i}$  (3) であり，

$Z_i$ は一般に正規分布に従う。

以下では、対象分野を女性学とし、 $T$ を“gender”，他分野コーパスの数を3とした場合 ( $1 \leq i \leq 3$ ) に、下記の値 (各コーパスにおける  $T$  の出現回数およびコーパスの用語総数) を設定し、計算例を示す。

- $W_0 = 20000$ ,  $f_0(\text{gender}) = 50$
- $W_1 = 19000$ ,  $f_1(\text{gender}) = 5$
- $W_2 = 21000$ ,  $f_2(\text{gender}) = 30$
- $W_3 = 22000$ ,  $f_3(\text{gender}) = 0$

$$i = 1: \frac{\frac{50}{20000} - \frac{5}{19000}}{\sqrt{\left(\frac{55}{39000}\right)\left(1 - \frac{55}{39000}\right)\left(\frac{1}{20000} + \frac{1}{19000}\right)}} = 5.88$$

$$i = 2: \frac{\frac{50}{20000} - \frac{30}{21000}}{\sqrt{\left(\frac{80}{41000}\right)\left(1 - \frac{80}{41000}\right)\left(\frac{1}{20000} + \frac{1}{21000}\right)}} = 2.46$$

$$i = 3: \frac{\frac{50}{20000} - \frac{0}{22000}}{\sqrt{\left(\frac{50}{42000}\right)\left(1 - \frac{50}{42000}\right)\left(\frac{1}{20000} + \frac{1}{22000}\right)}} = 7.42$$

上記のうち最小となるのは、 $i = 2$  の場合であり、 $MDP(\text{gender})$ は、2.46 となる。

同様に、 $T$ を“research”として下記の値 (各コーパスの用語総数は前述の例と同様とする) を設定した計算例も示す。

- $f_0(\text{research}) = 30$
- $f_1(\text{research}) = 50$
- $f_2(\text{research}) = 100$
- $f_3(\text{research}) = 20$

$$i = 1: \frac{\frac{30}{20000} - \frac{50}{19000}}{\sqrt{\left(\frac{80}{39000}\right)\left(1 - \frac{80}{39000}\right)\left(\frac{1}{20000} + \frac{1}{19000}\right)}} = -2.47$$

$$i = 2: \frac{\frac{30}{20000} - \frac{100}{21000}}{\sqrt{\left(\frac{130}{41000}\right)\left(1 - \frac{130}{41000}\right)\left(\frac{1}{20000} + \frac{1}{21000}\right)}} = -5.87$$

$$i = 3: \frac{\frac{30}{20000} - \frac{20}{22000}}{\sqrt{\left(\frac{50}{42000}\right)\left(1 - \frac{50}{42000}\right)\left(\frac{1}{20000} + \frac{1}{22000}\right)}} = 1.75$$

上記のうち最小となるのは、 $i = 2$  の場合であり、 $MDP(\text{research})$ は、-5.87 となる。従って research よりも gender の方が高い  $MDP$  の値を持ち、その意味で gender の方が research より女性学用語の可能性が高いと言える。

$MDP$  の値として最小値を採用している理由は、専門用語は、その用語が専門用語である分野以外の分野ではあまり多く出現しないという専門用語の性質を前提としているためである。すなわち、他分野のうち、たとえ1つの分野だけであったとしても、 $T$  の他分野コーパスでの出現率が高かったならば、 $T$  は専門用語ではない可能性がある。前述の例では、gender は他の3分野いずれにおける出現率よりも WSIF での出現率の方がずっと高いのに対して、research は他の2分野において WSIF での出現率よりもそれほど高くはない。従って、gender は research よりも、女性学のみ頻出する語である可能性が高く、その意味で gender の方が research よりも女性学の専門用語である可能性が高い。

## 4.2 比較対象手法

以下ではまず、 $UC$  について説明し、続いて  $C\text{-value}$ 、 $FLR$  について説明する。

### 4.2.1 $UC$

$UC$  は以下のように定義される[26].

$$UC(T) = f_0(T) \times \frac{f_0(T)}{DF} \times \left( \frac{1}{FF_i} \right)^3 \quad (4)$$

ここで：

$f_0(T)$ ：候補用語  $T$  の WSIF コーパスでの出現頻度

$DF$ ：対象コーパスでの  $T$  の出現する文書数

$FF_i$ ：全コーパス（対象分野および他分野コーパス）のうち、 $T$  が出現した分野の数

である。

以下では、 $MDP$  の計算例で使した値に下記の通り  $DF$  の値を加えて、**gender** と **research** の  $UC$  の値を求める。

- $f_0(\text{gender}) = 50, \quad DF = 30$
- $f_1(\text{gender}) = 5, \quad DF = 2$
- $f_2(\text{gender}) = 30, \quad DF = 10$
- $f_3(\text{gender}) = 0, \quad DF = 0$
- $f_0(\text{research}) = 30, \quad DF = 15$
- $f_1(\text{research}) = 50, \quad DF = 25$
- $f_2(\text{research}) = 100, \quad DF = 60$
- $f_3(\text{research}) = 20, \quad DF = 5$

$$UC(\text{gender}) = 50 \cdot \frac{50}{30} \cdot \left( \frac{1}{3} \right)^3 = 3.09$$

$$UC(\text{research}) = 30 \cdot \frac{30}{15} \cdot \left( \frac{1}{4} \right)^3 = 0.94$$

上記より、**gender** と **research** の計算結果を比較すると、 $UC$  の値が高い **gender** の方が、女性学の専門用語らしいと判断する。

#### 4.2.2 C-value

$C\text{-value}$  は以下のように定義される。

$$C\text{-value}(T) = \log_2 \left| T \left( f(T) - \frac{t(T)}{c(T)} \right) \right| \quad (5)$$

ここで：

$|T|$ ：候補用語  $T$  を構成している単語の数

$f(T)$ ：候補用語  $T$  がコーパス中に出現した回数

$t(T)$ ： $T$  を含んだ、より長い候補用語の延べ語数

$c(T)$ ： $T$  を含んだ、より長い候補用語の異なり語数

である。

以下では、“**gender**”および“**gender role**”を候補用語  $T$  とした場合に、これらを含んだ用語とその出現頻度（括弧内の数字）を示し、計算例を示す。

- brain gender role (2)
- driver gender role (1)
- gender role demand (3)
- gender role (50)
- gender identity (10)
- brain gender (6)
- driver gender (3)
- gender (100)

$$C\text{-value}(\text{gender}) = \log_2 1 \left( 175 - \frac{75}{7} \right) = 0$$

$$C\text{-value}(\text{gender role}) = \log_2 2 \left( 56 - \frac{6}{3} \right) = 54$$

上記の **gender** のように、1 語で構成された候補用語の  $C\text{-value}$  の値はすべて 0 となる。**gender** と **gender role** の計算結果を比較すると、 $C\text{-value}$  では **gender role** の方が女性学の専門用語らしいと判断する。

### 4.2.3 FLR

FLR は以下のように定義される.

$$FLR(T) = f'(T) \left( \prod_{i=1}^{|T|} (FL(t_i) + 1)(FR(t_i) + 1) \right)^{\frac{1}{2|T|}} \quad (6)$$

ここで:

$f'(T)$ : 候補用語  $T$  が単独 (より長い他の候補用語に含まれることのない形) で出現した回数

$t_i$ :  $T$  を構成する  $i$  番目の単語

$FL(t_i)$ :  $t_i$  の左側に隣接した名詞の延べ語数

$FR(t_i)$ :  $t_i$  の右側に隣接した名詞の延べ語数

である.

以下では,  $C$ -value で用いた例を使って  $FLR$  の計算例を示す.

$$FLR(\text{gender}) = 100 \cdot ((12+1) \cdot (66+1))^{\frac{1}{2}} = 2951.27$$

$$FLR(\text{gender role}) = 50 \cdot ((12+1) \cdot (66+1) \cdot (56+1) \cdot (3+1))^{\frac{1}{4}} = 2110.92$$

上記より,  $FLR$  では  $\text{gender}$  の方が  $\text{gender role}$  に比べて高い値を取ため,  $\text{gender}$  の方がより女性学用語らしいと判断する.

## 5 実験結果

### 5.1 提案手法と既存手法との精度の比較

各手法と, 2 つの正解用語集合を用いて, 精度・再現率を算出した. 本研究では以下のように, 精度・再現率を定義する.

- 精度 = (候補用語として抽出され, かつその候補用語が正解集合に含まれている用語数) / (候補用語として抽出された用語数)
- 再現率 = (候補用語として抽出され, かつその候補用語が正解集合に含まれている用語数) / (正解集合に含まれる用語数)

以下では, 各尺度のスコアが高い順に抽出した場合の精度・再現率を示す.

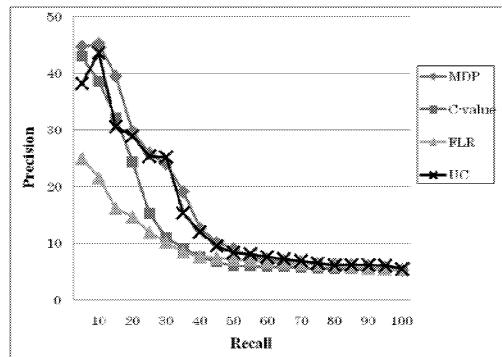


図1 正解集合 M に基づく各手法の精度・再現率

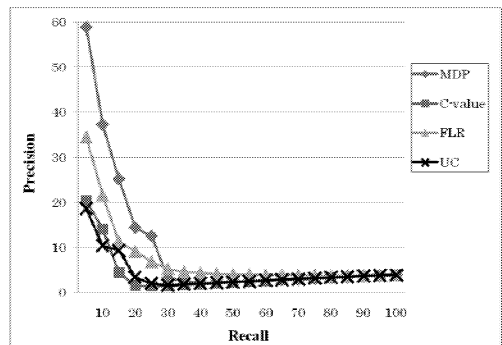


図2 正解集合 D に基づく各手法の精度・再現率

図1・図2にそれぞれ正解集合 M, 正解集合 D を用いた場合の各手法の精度・再現率を示す。図1より, 再現率が 25%以下の場合において, *MDP* は他の手法に比べ最も高い精度を示している。特に再現率 20%以下の場合には, *UC* よりもかなり高い精度を示した。図2より, 再現率 25%以下において, 明らかに *MDP* がその他の 3 手法よりも高い精度を示している。このことから, 本研究手法は従来の代表的抽出手法より高い精度を実現するという意味で, これまでにない効果的な手法であると言える。

図1から, *MDP*, *UC* の 2 手法の精度は, *C-value*, *FLR* よりも高いことがわかる。前述したとおり, *MDP*, *UC* の 2 手法は, 他分野コーパスを用いているのに対し, *C-value* と *FLR* は対象分野のコーパスしか用いることができない。この結果から, 対象分野以外のコーパスを用いることは, 専門用語抽出の精度向上に有効であると言える。

図1と図2を比較すると, *MDP*, *C-value*, *UC* の精度は正解集合 M よりも正解集合 D において低いことが分かる。一方で, *FLR* の精度は明らかに低くなっているとは言えない。また, *C-value* は, 他の尺度と比較すると, 上記の精度の差が顕著であるが, これらは次のように説明することができる。即ち, 正解集合 M に比べて正解集合 D は, より多くの (a) 1 つの語のみで構成される用語, (b) 複合語の構成用語となっている語 (“state”など), (c) 複数の分野に多く出現する用語 (“article”など), を含んでいる。*MDP*, *UC* は, 専門分野コーパスに多く出現し, 他分野のコーパスにはそれほど多く出現しない用語を抽出するために, (c) のように複数の分野において出現頻度の高い用語を抽出するのは難しい。また, 前述のとおり *C-value* は, 1 つの語から成

る用語には 0 という低いスコアを付与するため, 1 つの語のみで構成される (a) と, 大半は 1 つの語から成る複合語の構成用語である (b) は, スコアが低いため専門用語として抽出するのは難しい。一方で, *FLR* は, (b) を適切に抽出しやすい性質を持っているため, 他の尺度に比べて, 精度にあまり差がなかったと考えられる。このように, 図1と図2の違いは, 各尺度の性質によって説明することができる。

## 5.2 各手法における抽出用語の傾向

表1に, 各手法の専門用語抽出尺度の値が高い順に, 25 位までの用語を掲載した[27]。前述したように, 1 語で構成された候補用語の *C-value* の値はすべて 0 となる。そのため, 抽出スコアの上位にはこのような語は現れず, すべて複合語が掲載されている。

各手法それぞれの上位 25 位を見ると, 大きく分けて 2 種類の抽出誤りの傾向が見られた。一つは, 学術論文に特有な用語を抽出してしまうことである。この傾向は, 他分野コーパスとの比較を行っていない *FLR* (study, article, paper など) および *C-value* (case study など) で顕著に見られる。提案手法では, “article” がこれに該当する。*UC* では特に該当するものはない。

もう一つの傾向は一般的に使用される用語を抽出してしまうことである。これは本手法では “self” などが, *C-value* では, “United States” などが, *FLR* では, “work” などが, *UC* では, “Mari” (固有名詞) などが該当する。このように抽出スコアが上位の用語であっても, いくつかの抽出誤りが検出された。

表 1 4 手法による専門用語尺度の値上位 25 位まで

rank	MDP	C-value	FLR	UC
1	woman	feminist theory	woman	feminist
2	feminist	violence against woman	feminist	hijab
3	gender	United States	study	motherhood
4	feminism	higher education	gender	masculinity
5	self	feminist research	article	rape
6	movement	young woman	research	patriarchy
7	article	human right	work	Mari
8	feminist theory	black woman	paper	BFM
9	way	group of woman	state	feminist perspective
10	lesbian	domestic violence	experience	feminist research
11	struggle	feminist perspective	group of woman	peace movement
12	young woman	South Africa	movement	violence against woman
13	motherhood	world war	role of woman	feminist objectivity
14	violence	American woman	world	meaning of home
15	identity	case study	class	cosmetic surgery
16	femininity	number of woman	family	migrant woman
17	rape	same time	experience of woman	study student
18	body	role of woman	young woman	symbolic violence
19	patriarchy	division of labour	American woman	gender justice
20	masculinity	Muslim woman	status of woman	virginity
21	domestic violence	labour market	position of woman	other woman
22	feminist perspective	many woman	feminist research	class woman
23	Irish woman	sexual harassment	right	feminist methodology
24	American woman	child care	feminist theory	feminist pedagogy
25	academy	gender equality	man	prostitution

### 5.3 他分野コーパスの規模の検討

コーパス作成のコストを考慮すると、コーパスの規模は小さければ小さいほど良い。ただし、コーパスの規模を小さくすることで抽出精度も大幅に下がるということが無いよう何らかの工夫は必要である。ここでは実験を通して、コーパス規模の縮小可能性について検討する。

#### 5.3.1 コーパス規模を変えた精度の比較

ここで他分野コーパスを構成する全 39 ジャーナルのうち、任意の 3 ジャーナルによって作られるコーパスを新たな他分野コーパス (3 分野コーパス) とする。同様に、任意の 2 ジャーナルによって作られるコーパスを 2 分野コーパスとし、さらに任意の

1 ジャーナルによって作られるコーパスを 1 分野コーパスとする。

3 分野コーパスの場合、39 ジャーナルから 3 ジャーナルの組み合わせを計算すると、 ${}_{39}C_3=9139$  となる。そこで 9,139 通りの 3 分野コーパスを他分野コーパスとして、女性学用語の自動抽出を行った。この実験では正解用語  $M$  を用いて、精度・再現率を算出した。その結果、再現率 25%, 50%, 75% 時点の平均精度が高かったジャーナルの組み合わせは、

(1) *Educational Gerontology & International Journal of Public Administration & Urban Studies*, (2) *International Journal of Public Administration & Journal of Applied Statistics*



& Urban Studies, (3) *International Journal of Public Administration & Medical Teacher & Urban Studies*, の順であった。

同様の方法で、2 分野コーパスを他分野コーパスとした 741 通りの組み合わせのうち、再現率 25%, 50%, 75% 時点の平均精度が最も精度が高かったジャーナルの組み合わせは、(1) *International Journal of Public Administration & Urban Studies*, (2) *International Journal of Public Administration & Medical Teacher*, (3) *Educational Gerontology & Urban Studies*, の順であった。

また同様に、1 分野コーパス場合で実験を行ったところ、(1) *International Journal of Public Administration*, (2) *Urban Studies*, (3) *Educational Gerontology*, の順に平均精度が高かった。

図 3 に、39 分野全て、3 分野、2 分野および 1 分野の最も精度が高い組み合わせ、最も低い組み合わせの合計 7 パターンの精度・再現率を示した。

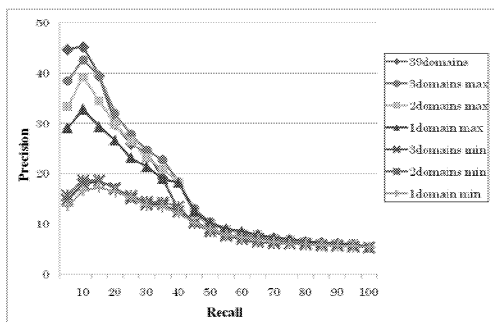


図 3 コーパス規模を変えた精度・再現率

図 3 から、再現率 15% 以下では、39 分野全てのコーパスを用いた場合 (39domains) が最も精度が高く、次いで 3 分野の最も精度の高い組み合わせ (3domains max), 2 分野の最も精度の高い組み合わせ (2domains

max), 最も精度の高い 1 分野 (1domain max) のコーパスを用いた順に精度が高いことがわかった。

さらに、これら精度の高い 4 つのパターン (39 domains, 3domains max, 2domains max, 1domain max) と、他の既存手法 (UC, C-value, FLR) の精度・再現率を比較したものを図 4 に示す。

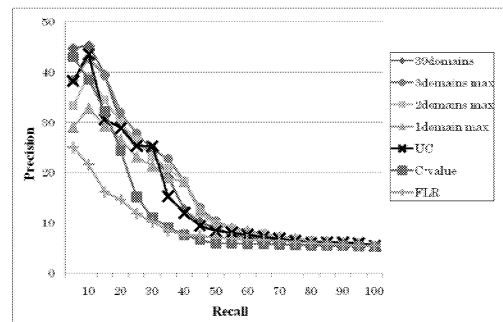


図 4 既存手法とコーパス規模を変えた精度・再現率

図 4 より、再現率 5% 以下および 15% 以上 25% 以下では、3domains max が他の既存手法よりも精度が高いことがわかった。このことから、39 分野全てをコーパスとして用いた場合には及ばないが、精度が高くなるような分野 (ジャーナル) の組み合わせを用いれば、コーパス規模を縮小しても、前述の再現率において既存手法よりも高い精度で専門用語抽出を行えることがわかった[28]。

### 5.3.2 WSIF コーパスと他分野コーパスの類似度

前項の結果から、行政学 (*International Journal of Public Administration*) や都市研究 (*Urban Studies*) のジャーナルを含んだコーパスを用いた時の精度が高いことがわか

った。逆に、神経科学 (*International Journal of Neuroscience*) や、生殖医療におけるシステム生物学 (*Systems Biology in Reproductive Medicine*) は精度が低かった。神経科学やシステム生物学は、自然科学分野に属した分野であるが、行政学、都市研究は、人文・社会科学分野に属した分野である。本研究の対象分野である女性学も後者の人文・社会科学分野に属している。このことから、コーパスの学問分野の類似度が高いほど精度が高くなると推測できる。そこで、各コーパスの類似度を計算し、女性学と類似度の高いコーパスが、前述の精度の高い組み合わせと一致するかを調べた。ここではベクトル空間モデルを用いて、WSIF コーパスと他分野コーパスの類似度を計算した。他分野コーパス  $C_j$  ( $1 \leq j \leq 39$ ) と WSIF コーパス  $W$  の類似度  $sim$  は、以下の式で表せる。

$$sim(C_j, W) = \frac{\bar{C}_j \cdot \bar{W}}{\|\bar{C}_j\| \|\bar{W}\|} \quad (7)$$

$$= \frac{\sum_{i=1}^m (c_j^i \cdot w_i)}{\sqrt{\sum_{i=1}^m (c_j^i)^2 \cdot \sum_{i=1}^m (w_i)^2}} \quad (8)$$

ここで：

$$\bar{C}_j = (c_j^1, c_j^2, \dots, c_j^m)$$

$$\bar{W} = (w_1, w_2, \dots, w_m)$$

$m$  : WSIF と他分野コーパスの 2 コーパスにおける候補用語の総異なり語数

$c_j^i$  :  $j$  番目の他分野コーパスにおける  $i$  番目の候補用語

$w_i$  : WSIF コーパスにおける  $i$  番目の候補用語

と定義する。

類似度の計算の結果、WSIF コーパスとの類似度が高かったジャーナル上位 3 件は、

1. *International Journal of Public Administration*
2. *Urban Studies*
3. *Educational Gerontology*

であった。この結果は、前述の精度の高いジャーナルの組み合わせと一致している。

すなわち、3 分野コーパスのうち最も精度が高かった “*Educational Gerontology & International Journal of Public Administration & Urban Studies*” の組み合わせ、2 分野コーパスのうち最も精度が高かった

“*International Journal of Public Administration & Urban Studies*”，および 1 分野コーパスで精度が高かった “*International Journal of Public Administration*” “*Urban Studies*” “*Educational Gerontology*” である。

このことから、女性学と類似した学問分野のコーパスを用いることで、精度をそれほど下げずに用語抽出が行えると言える。

## 6. 考察

提案手法と既存手法との精度の比較においては、正解集合  $M$  を用いて実験を行った場合に、対象分野以外のコーパスを用いた 2 手法 ( $MDP$  および  $UC$ ) の精度が高かった。このような比較は Uchimoto et al.[1]において行われておらず、他の研究でも示されていない。他分野コーパスを用いることが有用であるということを、先行手法と比較した実験により実証したことは、本研究が提示する新たな知見であると言える。

また、*MDP* の抽出誤りの原因については、*WSIF* コーパスと他分野コーパスでの出現率の差だけを考慮するため、ある用語が、他分野コーパスに高頻度で出現していたとしても、*WSIF* コーパスにその頻度以上で出現していた場合は、出現率に差があるとして、その用語を抽出してしまうことが考えられる。この問題に対しては、他分野コーパスでの出現頻度が少ない用語のみを抽出することが考えられる。出現頻度に関しては、適切な閾値を求める必要があるが、この部分については今後の課題としたい。

コーパス規模の縮小可能性については、実験を通して、対象分野と類似した学問分野のコーパスを用いれば、大きく精度を下げることなく、3 分野のテキストを用いたコーパスに縮小できることがわかった。このことは、既存の手法との比較からも明らかである。類似した学問分野のコーパスを使うことによって、高い精度で専門用語抽出が可能となるのは、分野共通用語とも言えるべき用語の存在が関係すると考えられる。

ここで言う分野共通用語とは、ある学問分野（たとえば、女性学や航空力学など）の専門用語ではないが、一般語でもない学術的な用語であり、人文科学分野などの広い概念における学問分野において共通する用語のことである。

図 5 では、例として自然科学分野と人文・社会科学分野といった異なる分野に属した学問分野（航空力学と女性学）における分野共通用語と専門用語の例を示す。一方、図 6 では、同じ人文・社会科学分野に属した学問分野（行政学と女性学）における分野共通用語と専門用語の例を示す。図 5・図 6 共に、内側の円は学問分野名とその分野の専門用語の例を示し、外側の円は分

野共通用語の例を示している。

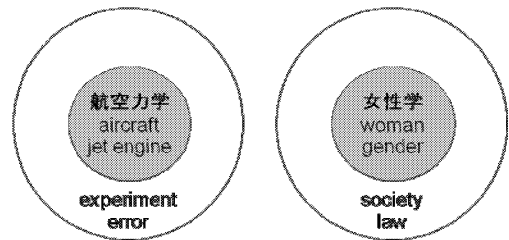


図 5 自然科学分野および人文・社会科学分野における分野共通用語と専門用語

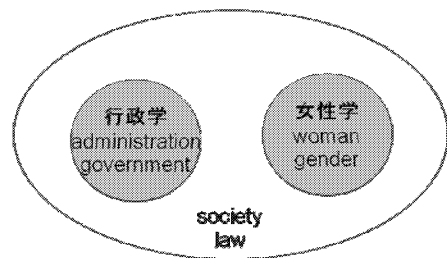


図 6 人文・社会科学分野における分野共通用語と専門用語

*MDP* は、2 つの分野コーパスにおけるある語の出現率の差を考慮する。図 5 の女性学と航空力学のように分野共通用語が異なる場合、“society”や“law”は女性学の専門用語ではないにも関わらず、航空力学のコーパスにこれらの語が出現しないため、*MDP* の値は高くなり、女性学の専門用語として抽出されてしまう可能性が高い。一方、図 6 の女性学と行政学のように、同じ人文・社会科学分野の場合は、行政学のコーパス中に“woman”や“gender”といった女性学の専門用語そのものが高頻度で出現する可能性は低い、 “society”や“law”などの分野共通用語が出現する可能性は高い。よって、“society”や“law”の *MDP* の値は低くなり、これらの分野共通用語が女性学の専門用語

として抽出されることはない。その結果、女性学と同じ分野共通用語を持った分野コーパスを用いることで専門用語を正しく抽出することができ、精度が高くなると考えられる。

しかし女性学の一分野のテキストを比較対象コーパスとして用いることは当然できないように、「類似した分野のテキスト」についてどの程度の類似度まで許されるのかは本研究で明らかにしていない。これについては今後の課題としたい。

## 7. おわりに

本研究の目的は、対象分野コーパスと他分野コーパスでの候補用語の出現比率の差を考慮に入れた新しい専門用語抽出手法を提案し、従来の専門用語抽出手法と比較することで、提案手法の有効性を示すことであった。実験の結果、(1) 本研究手法は先行研究手法のいずれよりも精度・再現率が高いこと、(2) 対象分野コーパスのみを利用する *C-value* と *FLR* よりも、対象分野以外の他分野コーパスを用いる本研究手法と *UC* の方が、精度・再現率が高いこと、がわかった。さらに、本研究では 39 の分野を比較対象として用いたが、対象分野と類似した分野のみを用いることで、コーパス作成のコストを減少させながら抽出精度・再現率をある程度維持できることを示した。

## 謝辞

女性学用語の専門家として、多くの助言および正解集合の作成に貢献していただいた独立行政法人 国立女性教育会館 情報課 専門職員の森未知氏（女性情報研究会メン

バー）と赤嶺良子氏に深く感謝すると共に、本稿の執筆にあたり、多くの助言をいただいた筑波大学大学院図書館情報メディア研究科 阪口哲男准教授に深く感謝する。

## 参考文献

- [1] Uchimoto, Kiyotaka; Sekine, Satoshi; Murata, Masaki; Ozaku, Hiromi; Isahara, Hitoshi: “Term Recognition Using Corpora from Different Fields”, *Terminology*, Vol.6, No.2, pp.233–256, 2000.
- [2] Anselmo, Peñas; Felisa, Verdejo; Julio, Gonzalo: “Corpus-Based Terminology Extraction Applied to Information Access”, *Proceedings of the Corpus Linguistics 2001 (CL 2001)*, Lancaster, pp.458–465, 2001.
- [3] 長尾真; 水谷幹男; 池田浩之: 「日本語文献における重要語の自動抽出」, 情報処理, Vol.17, No.2, pp. 110– 117, 1976.
- [4] Harter, Stephen. P: “A Probabilistic Approach to Automatic Keyword Indexing – Part I. On the Distribution of Specialty Words in a Technical Literature”, *Journal of the American Society for Information Science*, Vol.26, No.4, pp.197-206, 1975.
- [5] Harter, Stephen. P: “A Probabilistic Approach to Automatic Keyword Indexing – Part II. An Algorithm for Probabilistic Indexing”, *Journal of the American Society for Information Science*, Vol.26, No.5, pp.280-289, 1975.
- [6] Kageura, Kyo; Umino, Bin: “Methods of Automatic Term Recognition”, *Terminology*, Vol.3, No.2, pp.259-289, 1996.
- [7] 小山照夫; 影浦峽; 竹内孔一: 「日本語専門分野テキストコーパスからの複合語用語の抽出」, 情報処理学会研究報告. 自然言

語処理研究会報告, Vol.2006, No.124, pp. 55- 60, 2006.

[8] 山田寛康; 工藤拓; 松本裕治: 「単語の部分文字列を考慮した専門用語抽出と分類」, 情報処理学会研究報告. 自然言語処理研究会報告, Vol.2000, No.107, pp. 77- 84, 2000.

[9] Justeson, John. S; Katz, Slava. M: “Technical Terminology: Some Linguistic Properties and an Algorithm for Identification in Text”, *Natural Language Engineering*, Vol.1, No.1, pp.9-27, 1995.

[10] Ananiadou, Sophia: “A Methodology for Automatic Term Recognition”, *Proceedings of the 15th International Conference on Computational Linguistics (COLING 1994)*, Kyoto, pp.1034-1038, 1994.

[11] Steinacker, Ivo: “Indexing and Automatic Significance Analysis”, *Journal of the American Society for Information Science*, Vol.25, No.4, pp.237-241, 1974.

[12] Jones, Leslie. P; Gassie, Jr., Edward. W; Radhakrishnan, Sridhar: “INDEX: The Statistical Basis for an Automatic Conceptual Phrase-Indexing System”, *Journal of the American Society for Information Science*, Vol.41, No.2, pp.87-97, 1990.

[13] Cohen, Jonathan. D: “Highlights: Language-and Domain-Independent Automatic Indexing Terms for Abstracting”, *Journal of the American Society for Information Science*, Vol.46, No.3, pp.162-174, 1995.

[14] Nenadic, Goran; Ananiadou, Sophia; McNaught, John: “Enhancing Automatic Term Recognition through Recognition of Variation”, *Proceedings of the 20th International Conference on Computational Linguistics*

(*COLING 2004*), Geneva, pp.604-610, 2004.

[15] 中川裕志; 湯本紘彰; 森辰則: 「出現頻度と連接頻度に基づく専門用語抽出」, 自然言語処理, Vol.10, No.1, pp. 27- 45, 2003.

[16] Frantzi, Katerina; Ananiadou, Sophia; Mima Hideki: “Automatic Recognition of Multi-word Terms: the C-value/NC-value Method”, *International Journal on Digital Libraries*, Vol.3, No.2, pp.115-130, 2000.

[17] ここでは便宜上尺度の名称を UC とした.

[18] <http://www.sciencedirect.com/science/journal/02775395> から入手.

[19] Taylor & Francis (<http://www.taylorandfrancisgroup.com/>) から入手. ジャーナルの各号の抄録の一括ダウンロードが可能であり, 容易に収集出来たため利用した.

[20] <http://www.cs.jhu.edu/~brill/>

[21] 辞書の見出し語がその分野の専門用語を過不足なく含んでいるという保証はない. そのため, コーパスから手作業で正解用語を抽出し, 2 種類の正解用語を用いた. 正解集合作成のためのコストについては, 正解集合 M は 3 人でおおよそ 116 時間程度, 正解集合 D の作成は 1 人で行い, おおよそ 30 時間程度であった. なお, 正解集合 D のうち正解集合 M と一致する語は, 247 語(約 24%)であった.

[22] 1989 年版: Tierney, Helen., ed.: “Women's Studies Encyclopedia”, Vol.1, Westport, Conn., Greenwood Press, 417p, 1989.

1990 年版: Tierney, Helen., ed.: “Women's Studies Encyclopedia”, Vol.2, Westport, Conn., Greenwood Press, 381p, 1990.

1991 年版: Tierney, Helen., ed.: “Women's

Studies Encyclopedia”, Vol.3, Westport, Conn., Greenwood Press, 531p, 1991.

[23] 1999 年版 : Tierney, Helen., ed.: “Women's Studies Encyclopedia. Rev. and Expanded ed.”, Westport, Conn., Greenwood Press, 1,607p, 1999.

[24] *Women's Studies Encyclopedia* の見出し語の数は, 1,013 語 (1989 年版), 1,592 語 (1990 年版), 1,650 語 (1991 年版), 3,545 語 (1999 年版) であったが, そのうち人名と思われるものは除外し, さらに各巻で重複する用語を削除したところ, 3,312 語となった.

[25] 池田央 : 「統計ガイドブック」, 東京, 新曜社, 239p, 1989.

[26] ここでは, Uchimoto et al. が提案しているいくつかの尺度のうち, 対象分野コーパスと学問分野コーパスを比較しているもの

を取り上げ, さらに抽出精度において最高値を取ったパラメータを使用している.

[27] 4 手法による専門用語尺度の値上位 25 位までの用語のうち, 全手法で一致する用語はなかった. なお, *MDP* を含む 3 手法で一致した語は, American woman, feminist, feminist perspective, feminist theory, young woman であり, *MDP* と *UC* とで一致した語は, motherhood, masculinity, patriarchy, rape であった.

[28] テキストの収集やタグ付け・整形など, 専門用語自動抽出用コーパスの作成にはコストがかかる. コーパスの収録分野が例えば 39 分野でなく 3 分野で済んだなら, その作成コストは大幅に削減でき, 結果的に用語抽出全体のコスト削減につながる.

(2009年10月5日受付)  
(2010年1月13日採択)  
(2010年1月19日オンライン公開)

## 付録

本研究で使用したジャーナル名一覧

- ・ Aerosol Science and Technology
- ・ Applied Economics
- ・ Australian Journal of Earth Sciences
- ・ Avian Pathology
- ・ British Poultry Science
- ・ Clinical Toxicology
- ・ Current Eye Research
- ・ Early Child Development and Care
- ・ Educational Gerontology
- ・ Electric Power Components and Systems
- ・ Energy Sources, Part A: Recovery, Utilization, and Environmental Effects
- ・ Ergonomics
- ・ Ferroelectrics
- ・ International Journal of Computer Mathematics
- ・ International Journal of Control
- ・ International Journal of Electronics
- ・ International Journal of Mathematical Education in Science and Technology
- ・ International Journal of Neuroscience
- ・ International Journal of Polymeric Materials
- ・ International Journal of Public Administration
- ・ International Journal of Radiation Biology
- ・ International Journal of Remote Sensing
- ・ Journal of Applied Statistics
- ・ Journal of Clinical and Experimental Neuropsychology
- ・ Journal of Macromolecular Science, Part A
- ・ Journal of Modern Optics
- ・ Journal of Natural History
- ・ Journal of Statistical Computation and Simulation
- ・ Medical Mycology
- ・ Medical Teacher
- ・ Molecular Physics
- ・ Pathology
- ・ Petroleum Science and Technology
- ・ Phase Transitions
- ・ Radiation Effects and Defects in Solids
- ・ Systems Biology in Reproductive Medicine
- ・ The Journal of Adhesion
- ・ Urban Studies
- ・ Xenobiotica

調査報告

# 情報記録手法と記憶定着・理解度の関係についての実験報告 ～手書き記録時とキーボード記録時の差異について～

## A study on recording tools affecting memory and understanding ～A comparison between handwriting and keyboard input～

曾根原士郎\*, 齋藤敦子

Shiro SONEHARA\*, Atsuko SAITO

コクヨ株式会社

KOKUYO Co.,Ltd

〒108-8710 東京都港区港南 1-8-35

一般的に、手書きでの情報記録手段は、教育現場において学習・記憶・理解の効果を高めると言われている。しかし、近年のビジネス現場ではパソコンへのキーボード入力が主流であり、特に若いビジネスワーカーはそれらデジタルツールを巧みに使いこなしながら、効率的に業務を行っている。本実験では手書き記録手法の良さを再認識すると同時に、これからのビジネスワーカーがデジタルツールを用いてさらに生産性を高めていく可能性が示唆された。

In general, on educational situations it is said that handwriting will improve the effect of memory and understanding more than keyboard input. But recently, the keyboard input is the main current in business scenes and business workers skillfully master digital tool, so as to work efficiently. In this study we reconfirm the effect of handwriting, and suggest possibility that business workers in the future will improve the productivity with those digital tools.

キーワード: 手書き記録、キーボード入力記録、記憶、理解、ビジネススキル

Keyword: handwriting, keyboard input, memory, understanding, Business Skills

### 1. はじめに

これまでも手書き記録手法とキーボード入力記録手法の差異による記憶定着への影響に関する研究は各方面で行われてきたが、その多くは学童や学生を中心としたものであり、今後のビジネスマン、ビジネスシーン視点での変化の傾向、兆候に着目した実験はあまり例を見ない。しかし近年、オフィスのパソコン環境は1人1台が一般的となり、多くの業務がキーボード入力で行われる様になった事から、ワーカーのノートや手帳と言った手書き記録ツールの活用目的やシーン、期待

効果にも変化が生じつつあると考えられる。今回、入社8か月を経て基本的ビジネススキルと現場経験を習得した直後の社員を将来の対象ビジネスマンと想定して、記録手法別に各ビジネスシーンを想定した情報の記録と、その記憶定着率や理解度に差異が生じるのかを実験した。

### 2. 実験方法

#### 2.1. 実験参加者

入社8か月後の社員 20 名



## 2.2. 実験装置

Anoto デジタルペン・ペーパー、ノート PC(素材再生用、参加者利用用)、スピーカー(素材再生用 PC に接続)、タイマー、ビデオカメラ

## 2.3. 呈示刺激

第一部素材:参加者間に前提知識の差を生じさせない内容で、非常に単純な 5W1H の記憶を確認する事を目的とした、動物が主人公の架空の話(2 分程度)。第二部素材:業務シーンを想定し、構造化された話を記憶・確認する目的として、消防署をモデルとした組織構造と職務内容の説明(3 分程度)。第三部素材:業務上で最も頻繁に想定される打合せ議事の記録や他者から口頭指示、説明を受けるシーンでの記憶・内容理解の確認を目的とした、大学教授による企業の戦略要素や組織論についての講演(12 分程度)。

## 3. 手続き

参加者の日常業務でのノート利用頻度や IT リテラシー等を事前に確認し、そのスキル別に 3 つの会議室に振り分けた後、各部屋単位に「手書きでメモを取るチーム」(8 名:日常的にノート利用率が高い、以下、手書きチーム)、「PC でメモを取るチーム」(8 名:Powerpoint スキルが高い、以下、PC チーム)、「何も記録媒体を与えないチーム」(4 名、以下、媒体無しチーム)とした。

各素材は音声データ化し、各会議室で個別に再生用 PC で第一部から第三部まで順番に再生しながら参加者に聞かせた。その際、「手書きチーム」は利用者用 PC を起動せず、Anoto デジタルペンとペーパーを利用して内容を手書き記録させた。「PC チーム」は利用者用 PC 上に Powerpoint でメモファイル(白紙)を準備し内容をキーボード入力記録させた。「媒体無しチーム」は、ただ着席し利用者使用 PC も起動せず素材音声を聞かせた。

第三部設問終了後、干渉テストとして百マス計算問題(7 分)を紙で実施し、その後、全員

が利用者用 PC に準備された素材別設問付き解答ファイルを指示に従って順に起動し、解答を作成した。

第一部素材の設問は動物名や描写、5W1H に関する単純な単語記憶を確認する内容(第一部単語)を、第二部素材の設問は各組織の名称と職務の内容を答えさせる単語記憶(第二部単語)と組織を階層図で答えさせる構造確認(第二部構造)を、第三部素材の設問は講演内に含まれる重要語句にあたる単語抽出(第三部抽出)と各センテンス単位の内容理解を確認する記述式解答、及び上司への全体要旨の報告文作成(第三部理解)とした。その際、第一部・第二部の設問解答では一切メモや記録は見ずに行う事にし、第三部設問解答のみ各メモや記録を参照して良い事とした。('媒体無しチーム'ではそもそもメモ・記録を行わない為、第三部についても第一部・第二部と同様の条件とした。)

## 4. 実験結果

実験結果は次に示すようなものであった。

### 4.1. 設問別の正答率

記録手法チーム別の各設問解答の平均点を比較した結果、第一部設問解答では「媒体無しチーム」が最も成績が高く、以降は全て「PC チーム」の成績が高かった。

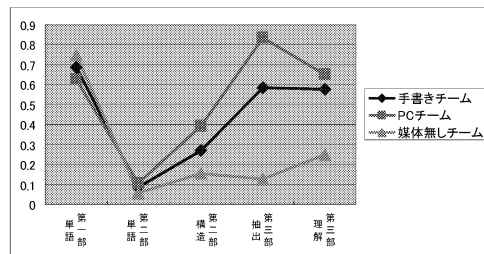


図 1:解答別の平均点推移

当初の仮説では第二部の構造確認や第三部理解確認の設問においては、「手書きチーム」が聞き取り内容を描画化する自由度(簡単に構造を図化して書き取る事ができる)の点から、記憶定着率的にも効率が良い手法として優位になると想定していたが、結果として「手書きチーム」の成績は第二部単語以降、

他の手法を上回る事がなかった(図1)。又、第三部単語抽出の設問では、「手書きチーム」と「PC チーム」はメモを見ながら回答するとした事によって、「媒体無しチーム」との差が大きかった。

#### 4.2. 平均値の差の検定

統計分析の結果、第三部設問の単語抽出設問解答において、「手書きチーム」と「PC チーム」との間に平均値の差の t 検定を行ったところ、 $t=-2.51(df=14)$ となり、5%水準で有意であった。

#### 4.3. 正解文字数の比較

「手書きチーム」と「PC チーム」のメモ記録(入力)文字数とその内に含まれた正解解答部分の文字数を比較した結果、第三部に大きな差が見られた(図 2)。しかし、同文字数をもその比率で比較した場合は逆に差異は見られなかった(図 3)。

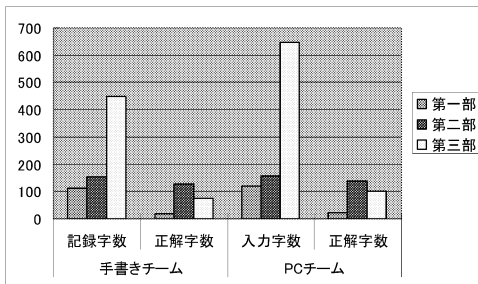


図 2:メモ記録文字数とその内の正解文字数比較

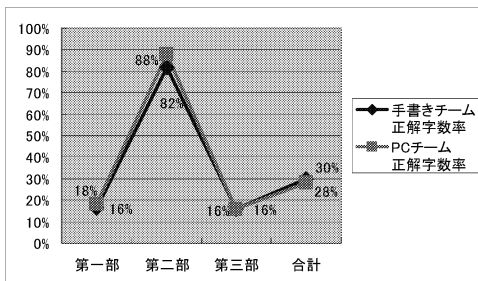


図 3:メモ記録文字数内の正解文字数の比率比較

#### 4.4. 得点のレンジ

「手書きチーム」と「PC チーム」の参加者個々の得点と各所属チームの平均点との差

を見た場合、「手書きチーム」では最高得点者と最低得点者の得点差は±6 点の範囲(図 4)にあるのに対して、「PC チーム」内での同得点差は±8.6 点となった(図 5)。又、「媒体無しチーム」での同得点差は±7.25 点であった。「媒体無しチーム」は元々参加者が 4 名と少ない上に、高得点と低得点に 2 名ずつ分れてしまっている為、データ、平均点共に参考程度とした。

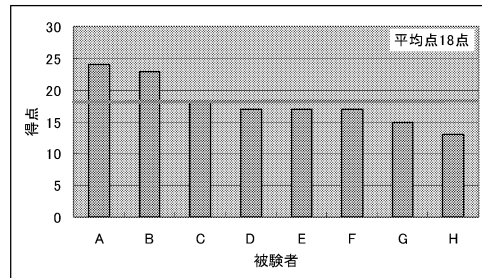


図 4:手書きチームの各自得点と平均点の差

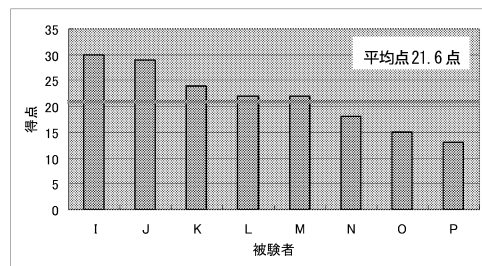


図 5:PC チームの各自得点と平均点の差

## 5. 考察

### 5.1. 単純な設問

図 1 の結果から、第一部素材では「媒体無しチーム」が記録ツールの干渉を受けず集中できた事により、記憶定着率が高まったと考えられる。元々第一部素材は本文 325 文字程度、内正解部分の字数は 40 文字程度で構成される単純なものだった。しかし「手書きチーム」は初めて手にする「Anoto デジタルペン・ペーパーと利用ルール」に、「PC チーム」は準備された「自分の物ではない PC」と「メモファイル(Powerpoint ファイル)」操作に意識が分散していたと想定でき、統制環境の差異と考えられる。

### 5.2. 複雑な設問

同じく図1の結果から、実験が進むにつれ「PC チーム」の集中力・記憶定着率が高まった可能性が考えられる。今回の参加者は入社1年次の一般社員であったにもかかわらず、「メモファイル(Powerpoint ファイル)」と同実験ビデオで確認したところ、「PC チーム」での第二部、第三部の平均文字入力速度は50~55字(変換後字数)/分と考えられ、ヒアリング・メモ入力中は十分に集中できていたと考えられる。

特に図2に見られる様に、第三部でのキーボード入力文字数は「手書きチーム」の1.4倍にも達しており、その集中状態での聴写力(聞き取り書き写す/入力する能力)が高まった事によって第二部構造や三部単語抽出の得点差が生じたと考えられる。

### 5.3. 記録文字数と解答文字数

しかし図3の結果からは、記録/入力文字数とその中に含まれる正解部分の文字数の比率は「手書きチーム」「PC チーム」の間で差が見られない。このことから、聴写力は両チーム間に大きな差がない、若しくは聴写力は記録手法に起因する記録量/スピードと関連し一定の比率範囲に止まる傾向があると考えられる。言い換えれば、より多く記録できるからと言って正解部分が飛躍的に多くなる事や、記録できる量が少ないからと言って的確に正解部分のみを記録する、という傾向は見られなかった。

### 5.4. 単語記憶の正解率

一方、同じ図3の結果から、第二部では「手書きチーム」、「PC チーム」共にメモ記録文字数に占める正解文字数率が非常に高いにもかかわらず、同単語記憶の設問正解率が「媒体無しチーム」と変わらない程度であった事は、聴写力と単語の記憶定着率の間に相関が無いとも考えられる。特にこの第二部素材は400文字弱のコンテンツであったが、所謂エピソード記憶化し難い組織名称や職務内容の羅列で構成されていた為、聴写力

に基づく単語記録は正確になされていても、その記銘や保持には至っていなかったと考えられる。

逆に同構造設問解答では「PC チーム」、「手書きチーム」共に「媒体無しチーム」を大きく上回った事から、聴写力による全体俯瞰された構造化と記憶定着率には相関性があると考えられる。

### 5.5. 手書きの優位性(1)

図1の結果から、第二部構造設問解答では「PC チーム」に、当初仮説立てしていた「手書きチーム」の優位性を上回る成果が見られた。この両チームの第二部メモを添削・検討したところ、「手書きチーム」のメモの多くには、予想通り円や線で階層を表現した自然な形での記録がみられた。しかし「PC チーム」のメモ(Powerpoint ファイル)では、改行やスペース挿入、タブ機能等を使って工夫/苦労しながら入力された形跡が多く見られ、それらの描画・操作ストレスが、逆に強い記憶定着をもたらした可能性が考えられる。

### 5.6. 入力スキル

図4の結果から、「手書きチーム」内で得点差が最も少なかった事は、「人の話を聞いて、必要事項をノートに書きとめる」と言う基本動作と記憶定着の訓練が同チーム内参加者間に一定のレベルで存在していたと考えられる。逆に図5で示した「PC チーム」内の下位得点者2名のキーボード入力字数は、残り6名の平均に対して57%程度であり、入力スピード・スキルに関する訓練不足、差がそのまま得点、記憶定着に影響を及ぼしていると考えられる。

### 5.7. 手書きの優位性(2)

「手書きチーム」と「PC チーム」の第三部抽出・理解の結果に限ってみた場合、それぞれの図1平均点推移と図2記録文字数を比較すると図6の様な関連が見られる。前述の通り第三部で両チームは記録した各自のメモを見ながら解答作成を行ったが、その記録文字

数をベースに平均点を見ると、「手書きチーム」が少ない記録文字数にもかかわらず効率的に得点を挙げたのに対し、「PC チーム」は記録文字数の多さが、内容理解解答や上司への報告書作成時に非効率性を招いたとも考えられる。

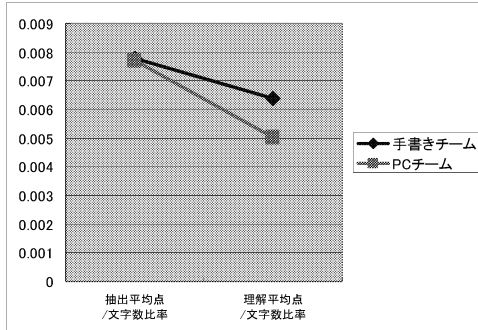


図 6:記録文字数に対する各平均点の比率

言い換えれば、均一なテキストによるメモ記録の中では重要度や強弱、要素間の関係性が見えず、見返しづらい構造となっていたが故に、解答作成に必要・適切な要素が探し出せない状況になってしまっていたと考えられる(図 7)。

会社の規模は利益収益性関係なし。	
内部管理コスト	差→利益収益
市場取引コスト	
すべての取引にはコストがかかるというのでノーベル賞	
1つ橋 警備員を警備会社に委託	
Why 内部管理コストのほうが高い	
大学の規模とその価値は関係ない	
組織の大きさと、質や利益は関係がない	
組織が大きくなるときは内部管理が肥大化するとき(市場取引)でまかなえない	逆も然り
BPO OS は内部管理コストの肥大化をビジネスチャンスとする	
コアコンピタンスさえあればあとは戦略的提携でよい会社経営がなされる。	
規模と価値を連動しないこと	

図 7:PC チーム得点上位者のメモファイル

逆に「手書きチーム」の得点上位者のメモでは、重要度や強弱に応じた文字サイズの書き分けや丸で囲む、矢印で要素間の関係を表す等の工夫が見られ、メモを見ただけでも、どの様な話の流れであったかが理解できるものとなっていた(図 8)。この点は考察 5.3、5.4.と合わせて考えると、手書きによる記録手法を使った想起・再生では、単純な文字と文字数だけでは無く、各種要素の識別や意図・ニュアンス、階層化が自由な描画・形状表現が可能で、非常に大きな効果・効用があると

考えられる。

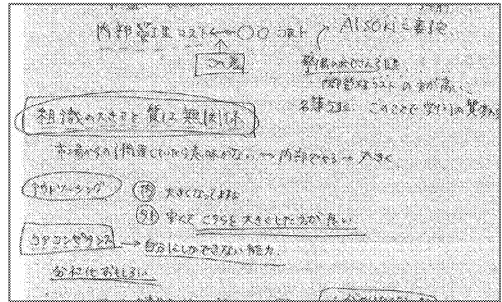


図 8:手書きチーム得点上位者のメモ

## 5.8. メモの重要性

第三部理解に関する設問解答(上司への報告文作成)内容を添削・検討した結果、「手書きチーム」、「PC チーム」では、同単語抽出設問解答に見られた差以上の質の差は見られなかった。この事は、ヒアリング時の聴写力による正確な各単語抽出能力と文意全体を把握・理解し表現する能力は別であると考えられると同時に、同設問解答において「媒体無しチーム」ではほとんど体をなさない状態であった事からも、ある程度正しい文章をまとめ上げる為には、その正確性を問わず、後から内容を想起・再生する為の何らかの意味ある情報を書き留める/入力する事が非常に重要だと考えられる。

## 6. まとめ

ビジネスシーンでは単純な 5W1H を中心とした情報にとどまらず、より高度で複雑な業務情報の口頭指示・説明の記録・記憶が正確且つ要点を押さえてスピーディーに行われる事が求められる。この必然性への対処方法として教育の現場で培われ訓練されてきた手書き記録手法とそれをベースとした記憶定着効果は疑うところではない。今回 PC を使った記録・記憶手法と合わせて、対象情報が複雑になるほど、記録媒体を使わない場合との記録・保持・想起の差は明確であり、現実のビジネスシーンにおいても、業務情報を「記憶にとどめる、理解する」為には「(手書き/PC 問わず)記録する」事は必須で

ある事が本実験結果からも見て取れる。

しかし同時に、今後の更なる PC の普及や業務上での活用環境の拡大を通じてリテラシー、スキルが強化されてゆく事で、PC を使った記録手法は、その記憶定着や内容理解・文章化効果の点で従来の手書き記録手法を上回る可能性が示唆されたとも考えられ、これ事は大きな発見であった。

今後は引き続き手書き手法の価値・効果の研究を進めると同時に、考察 5.6.、5.7. で述べた様な、手書き記録手法の持つ特性を生かしつつ効果的な記録・記憶(記銘・保持・想起)を支援する、PC 活用環境やシステムソフトウェアを含めた仮説を構築し、その実現方法や可能性、それらを用いた記録・記憶への効果等の調査を検討・考察してゆく。

## 7. 今後の課題

### (1)

各チームとも参加者数が少なく、事前にスクリーニングは行ったものの、参加者の特性に強く依存している可能性が高かったと思われる。今後、本実験を検討する場合は参加者数を増やすと同時に、同一人物への同質実験で検証する方法を検討したい。

### (2)

各種設問の内容と提示順序、時間配分、音源データの再生速度や音量に更なる統制・考慮が必要だったと感じる。又、今回実験中の統制(記録媒体個々の不具合等)に若干差が生じていた可能性があり、より正確な統制の統一を検討したい。

### (3)

各種設問設計と解答分析手法の精度を更に高める必要があると感じた。特にビジネスシーンを想定した「理解度」確認に関する記述解答とその解釈・分析について、次回は充分に検討の時間を組み立てたい。

### (4)

今後は記録媒体に、目的に対する効果を仮説立てして設計した意匠を伴った場合の手書き/キーボード入力記録手法の効果の差異含めた検証も検討したい。

## 謝辞

実験にご協力頂きました皆様に感謝致します。

## 参考文献

1. 吉村幸、他:『「書く」ことへの態度及び習慣と報告型作文の評価との関連について』、日本行動計量学会大会発表論文抄録集(37)、56-57、2009
2. 鈴木慶子、他:『視写と聴写に関する基礎調査(1)』、書写書道教育研究 23 号 11-20、2009
3. 大山典子、他:『デジタルノートテイキングに適した入力方法の研究 ～手書き入力とキーボード入力の差をめぐって～』、情報処理学会第 70 回全国大会講演論文集、“4-607”-“4-608”、2008
4. 上野賢太郎、他:『ノートテイキングにおける手書きとワープロの差異の研究～キーワードに注目した比較実験の視点から～』、情報処理学会第 70 回全国大会講演論文集、“4-609”-“4-610”、2008
5. 小川美沙都、他:『ノートテイキングにおける手書きとワープロの質的な差に関する検討(4)』～講義内容の記憶と視点に関する考察～、情報処理学会第 70 回全国大会講演論文集、“4-743”-“4-744”、2008

解説

## 著作物のフェアユースまたは公正使用に関する動向

### Fair Use of Copyrighted Materials

時実象一<sup>1\*</sup>,  
Soichi Tokizane<sup>1\*</sup>

1 愛知大学

Aichi University

〒441-8522 愛知県豊橋市町畑町 1-1

E-mail: tokizane@aichi-u.ac.jp

2010 年 1 月の文化庁文化審議会著作権分科会法制問題小委員会における報告「権利制限の一般規定について」において、限定された範囲で「権利制限の一般規定」、すなわちフェアユースを導入することが提言された。この報告と、2008 年 12 月の知的財産戦略本部会合における報告「デジタル・ネット時代における知財制度の在り方について」との関連について述べた。また海外のフェアユース規定について述べた。

A report to the Copyright Committee of the Agency for Cultural Affairs of Japan, in January 2010, recommend to introduce the fair use clauses only in a few selected areas. The relationship of this report with the 2008 report to the Council for Intellectual Property Promotion was discussed, as well as similar laws in other countries.

キーワード: 著作権, フェアユース, 公正使用, 権利制限, 文化庁, 知的財産戦略本部

Copyright, Fair Use, Limitation on exclusive rights, Agency for Cultural Affairs, Council for Intellectual Property Promotion

### 1 最近の動向

本年 (2010 年) 1 月 20 日に開催された文化庁文化審議会著作権分科会法制問題小委員会において、「権利制限の一般規定ワーキングチーム」から「権利制限の一般規定について」という報告がおこなわれた [1]。これは 123 ページにおよぶ大部のものである。なおここでいう「権利制限の一般規定」とはいわゆる「日本版フェアユース」のことである。なおフ

ェアユースは「公正使用」と訳される。

新聞等の報道によれば、今回の報告ではフェアユースの導入が可能とされる範囲は限定的で、

- (1) 広告で利用する写真にたまたま美術品などが写り込んでいるケース
- (2) 合法的な利用に必要なケース (CD をインターネット配信する場合のサーバーでの楽曲複製など)
- (3) 本来の利用でない複製 (言語分析のために小説を複写するなど)

に限られるとしている。要望のあったテレビ番組の転送サービスなどは適用対象から外れる見通しと伝えられている。

## 2 「権利制限の一般規定について」の報告の概要

この報告書では、権利制限を求められた利用行為を次のように分類している。

- ① いわゆる「形式的権利侵害行為」（利用の質または量が軽微であり実質的違法性がないと評価される行為）
- ② いわゆる「形式的権利侵害行為」と評価するか否かはともかく、その態様等に照らし権利者に特段の不利益を及ぼさないと考えられる利用
- ③ 既存の個別規定の解釈による解決可能性がある利用
- ④ 特定の利用目的を持つ利用
- ⑤ その他

これらについて文化庁ホームページの解説からまとめてみることにする。

### ① いわゆる「形式的権利侵害行為」

たとえば「写真や映像の撮影に伴ういわゆる「写り込み」」のように、付随的に著作物の利用が発生したもので、利用の程度が軽微であるものとされる。これについては「社会通念上著作権者の利益を不当に害しない」との意見が大勢であったと報告されている。

### ② その態様等に照らし権利者に特段の不利益を及ぼさないと考えられる利用

これに該当する場合としては、ひとつは、「適法な著作物の利用を達成する過程に

おいて不可避免的に生ずる」利用であり、「CDへの録音許諾を得た場合におけるマスターテープ等中間過程での複製」あるいは「著33条1項に基づく教科書掲載に関し、教科書作成過程での複製」等である。もうひとつは「著作物の表現を知覚（見る、聞く、読む等）するための利用とは評価されない」利用であり、「研究開発分野や情報の複製等を不可避免的に伴うネットワーク産業の分野」での利用とされている。これはたとえば言語分析のために小説を複写することなどが該当する。これらについても①と同様、「著作権者の利益を不当に害しない」との意見が大勢であったと報告されている。

### ③ 既存の個別規定の解釈による解決可能性のある利用

これについては、個別に解決可能なため、一般規定の対象とする必要はないとの意見が大勢であったが、「柔軟な適用が困難なものがあるため、要件を緩和（抽象化）する方向で見直しをすべき」との意見もあったとのことである。

### ④ 特定の利用目的を持つ利用

そのひとつは「障害者福祉、教育、研究、資料保存」など公益性のある利用であるが、これらについては必要に応じて個別規定を改正、創設すべきであるとの意見が大勢であった。またもうひとつは「パロディ」としての利用であるが、「表現の自由との関係、同一性保持権との関係等を慎重に検討する必要」があり、やはり必要に応じて個別規定を改正、創設すべきであるとの意見が大勢であったと報告されている。

⑤ 「その他」

この例としては、企業内の会議で書籍のコピーを配るなどの利用があるが、これらについても個別規定で対応すべきであるとされ、また番組録画転送サービスについては間接侵害の問題との意見となったと報告されている。

### 3 「デジタル・ネット時代における知財制度専門調査会」の報告

そもそもフェアユース問題の発端は、2008 年 12 月に麻生内閣における知的財産戦略本部会合において、中山信弘東京大学名誉教授を会長とする「デジタル・ネット時代における知財制度専門調査会」が提出した「デジタル・ネット時代における知財制度の在り方について」との報告 [2] である。この報告は「II. 権利制限の一般規定（日本版フェアユース規定）の導入」の「5. 検討結果」において次のように述べている。

「例えば、情報通信技術を活用した新しい産業の創出という観点からは、現行の著作権法では個別の制限規定が想定していない新規分野への技術開発や事業活動に対して萎縮効果を及ぼしているという問題がある。この点については、本専門調査会のヒアリングにおいて、事業者から同旨の意見があったほか、権利制限の一般規定は著作物の利用のルールを事後に決するというものであって、それを導入することにより、創造的な事業への挑戦を促進すべきという意見もあった。

また、ネット上の写真・動画への写り込みやウェブページ印刷などの行為には、形式的には違法となるが、権利者の利益を実質

的に害しているとは考えられず、また、社会通念上も違法とすべきとは考えられないものもある。一方、本専門調査会のヒアリングでは、権利者からは、一般規定の導入により違法な利用行為が蔓延するのではないかと、また、司法の判断によってしか解決できないこととなる結果、権利者に更なる負担を強いることになるのではないかとという意見があった。

以上のことから、個別の限定列举方式による権利制限規定に加え、権利者の利益を不当に害しないと認められる一定の範囲内で、公正な利用を包括的に許容し得る権利制限の一般規定（日本版フェアユース規定）を導入することが適当である。」

この報告書では、「形式的権利侵害」の具体例はほとんど示していない。わずかに示されている「写り込み」については、法制問題小委員会は前向きな意見であったのは前述のとおりである。次に示されている「ウェブページの印刷」については、日本文藝家協会、日本書籍出版協会、日本新聞協会など 6 団体が反対の意見書を文化審議会に提出している [3]。すなわちウェブページには有料のものも多く、単に「形式的権利侵害」とみなすのはおかしいとの意見である。

### 4 ヒアリング等でだされた主な意見

法制問題小委員会におけるヒアリング等で出された主な意見が報告書中にまとめられている。意見は非常に多岐にわたるが、権利制限の対象となりうる具体的事例が記載されており、参考となると思われるのでこれを付録 1 に示した。



## 5 海外におけるフェアユースの状況

### 5.1 米国の状況

よく知られているように、米国著作権法 107 条にはフェアユース規定がある [4]。それによれば、「著作物の使用がフェアユースとなるか否かを判断する場合に考慮すべき要素は、以下のものを含む」としており、次の4項目を挙げている。

- ①使用の目的および性質（使用が商業性を有するかまたは非営利的教育目的を含む）
- ②著作権のある著作物の性質
- ③著作権のある著作物全体との関連における使用された部分の量および実質性
- ④著作権のある著作物の潜在的市場または価値に対する使用の影響

これにより「著作者の許諾を得ずにリンクを張り利用者に提供すること」、「画像の縮小版であるこのサムネイル画像の表示」などが裁判を経てフェアユースであるとされており、その結果として検索エンジンでの画像検索結果の表示が可能となっている。特に「サムネイル」に関しては、職業写真家 Kelly が Arriba 社を告訴した裁判の第 9 巡回裁判所の判決において、判決は公正使用の 4 原則のうち 2 件 (①, ④) が最も重要であるとした [5]。まず使用と目的はアクセスを増すため、社会的要請に合致している。またこれにより、元の著作物の経済的利益が損なわれたとはいえず、むしろアクセス数を増やしたともいえるとしている。最近話題となっている Google Book Search の係争において、Google は当初この判決を引いて、無許諾での書籍の電子化はフェアユースである主張していた。しかし原告との和解がおこな

われたため、Google の行為がフェアユースであるかどうかについては判断が示されないままとなっている [6]。

### 5.2 その他の国の状況

前記法制問題小委員会の報告 [1] では、各国の調査の結果を次のように分類している（主なもののみ示す）。

表 1. 各国のフェアユース関連規定

規定の種類	国
米国型のフェアユース規定	米国、イスラエル、台湾、フィリピン、スリランカ
フェアディーリング規定	カナダ、英国
個別権利制限規定のみ	フランス、ドイツ、イタリア、スペイン、中国

ここで「フェアディーリング規定」とは「フェアユース」と類似しているが、個々の権利制限内容毎に定めているものである [7]。

## 6 今後の動向

今回の法制小委員会の報告は (1) フェアユース（権利制限の一般規定）の導入自体については前向きである、(2) ただしその内容については極めて限定的である、といった結果となり、表 1 にある多くのニーズに対しては答えられないままとなった。新しい民主党政権が著作物の円滑な利用とフェアユースについてどのような考えであるかについては明らかでない。むしろ鳩山首相や川端文科相が著作権保護期間の延長について述べるなど、逆行する発言もあるので [8]、今後の動向を注視したい。

## 参考文献

- [1] 文化庁文化審議会著作権分科会法制問題小委員会 平成 21 年第7回議事録. 2010/1/20.  
[http://www.bunka.go.jp/chosakuken/singikai/housei/h21\\_shiho\\_07/gijiyoshi.html](http://www.bunka.go.jp/chosakuken/singikai/housei/h21_shiho_07/gijiyoshi.html) (閲覧 2010/2/7).
- [2] 知的財産戦略本部デジタル・ネット時代における知財制度専門調査会. デジタル・ネット時代における知財制度の在り方について (報告). 2008/11/27.  
[http://www.ipr.go.jp/e\\_material/diginet\\_houkou.pdf](http://www.ipr.go.jp/e_material/diginet_houkou.pdf) (閲覧 2010/2/7).
- [3] 社団法人日本文藝家協会, 他. 「権利制限の一般規定」導入に関する意見書. 2010/1/20.  
<http://www.pressnet.or.jp/info/seimei/iken20100120.pdf> (閲覧 2010/2/7).
- [4] 高田寛. Web検索サービスの法的諸問題. 情報管理. 2009, 52(5), 267-275.
- [5] 時実象一. 国際図書館連盟 (International Federation of Library Associations: IFLA) 2006 年ソウル大会における「著作権とその他の法律問題」セッション. 情報の科学と技術. 2006, 56(11), 542-546.
- [6] 植村八潮. Googleブック検索プロジェクト--〈デジタル維新〉にどう立ち向かうか. 図書館雑誌 2009, 103(12), 820-821.
- [7] 北岡弘章. 日本版フェアユース規定の導入 [3] 英国のフェアディールングによる権利制限規定.  
<http://itpro.nikkeibp.co.jp/article/COLUMN/20081020/317314/> (閲覧 2010/2/7).
- [8] 時実象一. 著作権保護期間 延長は創作活動を阻害する. 朝日 2009/12/10 朝刊.

## 付録. 法制問題小委員会におけるヒアリング等で出された主な意見

区分	番号	内容	発言者
A	弱者者向けの拡大文字による一般図書など、公益目的での利用		延長問題フォーラム
B	視覚・聴覚障害以外にも多様な多様である。通常の著作物の利用が困難な人への利用や、緊急災害時などの障害者等への情報提供時における利用		障害者放送
C	点字を健常者にも分かるように文章に書き直す行為		インターネットサービス事業者
D	腐食の進んだ映画作品などの8ミリフィルムを保存のためにデジタル媒体に記録し直す行為など、文獻・映像保存などの各種アーカイブ		延長問題フォーラム、インターネットサービス事業者
E	権利者への経済的損害の極めてわずかな二次創作		延長問題フォーラム
F	企業内における著作物の利用		JEITA, MCF, 知財協
	① 情報共有や説明等のために企業内で著作物を利用する行為		
	1 外国語で書かれた論文、外国のウェブページを理解するために、従業員が翻訳すること		JEITA, 知財協, インターネットサービス事業者
	2 従業員がウェブサイトを参考文庫をダウンロードして印刷すること		JEITA, インターネットサービス事業者
	3 企業内の会議に用いるためにウェブサイトを新聞を転写すること		知財協, インターネットサービス事業者
	4 特許された技術の研究するために、他人の出願関係書類(特許明細書等)を複製すること		JEITA, インターネットサービス事業者
	5 特許行から受領した拒絶理由通知に添付された引用文献を、社内において拒絶理由を検討するために必要な範囲で複製すること		知財協
	6 営利目的の研究において全ての受講者がテキストとして市販本を購入し、講義の都合上、当該市販本の図をプロジェクトで撮影すること		JEITA
	7 購入した市販のマニュアルを、業務で使いやすいように必要な部分だけを複製してダイジェスト版を作成すること		JEITA
	8 正当に取得した図書を滅失に備えて、保管目的で電子的に複製すること		知財協
	9 遠方の会議に出張している社員からの問い合わせに回答するために、会社で購入した書籍の関連ページをFAXで送信する行為		インターネットサービス事業者
	10 社内会議で用いた資料を保存用に電子化(スキャン等)する行為		インターネットサービス事業者
	11 判例情報を社内で共有するために、訴訟の対象となった著作物の写真を転載すること		知財協
	12 裁判における準備書面、鑑定書、判例研究のため複製すること		知財協
	13 自身が取材を受けたテレビ番組を広報部門が録画保存する行為		インターネットサービス事業者
	14 社員がセミナーで講師の話を要領筆記する行為		インターネットサービス事業者
	15 業務上でやり取りする電子メールの転送の際に元のメールを複製する行為		インターネットサービス事業者
	16 US規格等技術標準規格をコピー、ダウンロードする行為		インターネットサービス事業者
	② 商品やサービスの説明、デモ目的の著作物の利用		MCF, 知財協
	17 サービスや企画内容を説明する目的で企画書や提案書における説明やデモ等に著作物を一時的に利用すること		MCF
	18 試験や立ち読みのように、商品購入の判断のために書籍の表紙・目次や音楽アルバムアート、内容の一部を利用すること		MCF
	19 レコード店が販売促進目的で販売中のCDを来店客に試聴させる行為		インターネットサービス事業者
	20 AV機器の販売店において、顧客に機器の性能を説明するために、CDから音楽を録音し、試聴させること		知財協
	21 展示会において、来場者に新発売のパソコンのブラウザ機能を説明するために、他人のウェブサイトをプロジェクトで映し出すこと		知財協
	22 家電量販店がテレビの高画質を来店客に訴求するために市販のDVDを店頭で再生する行為		インターネットサービス事業者
	23 家電量販店がオーディオの品質を来店客に訴求する目的で市販のODを再生する行為		インターネットサービス事業者
	③ 媒体変換のための著作物の複製		
	24 利用端末にあわせてファイル形式やサイズを変更すること		MCF

区分	番号	内容	知財協	発言者
	25	正當に取得した著作物について、当該著作物を再生できない他の種類の機器で再生できるように、異なる媒体に複製したり、ファイル形式を変換したりすること	知財協	
	26	ユビキタス目的のデータ形式変換	インターネットサービス事業者	
	27	HTMLをプログラムから利用するために文書をツリー構造にしたデータ形式であるDOM(Document Object Model)として利用する行為	インターネットサービス事業者	
	28	画像ファイルのファイル形式を変換(例: .jpg形式からgif形式)する行為	インターネットサービス事業者	
	④ 自己の著作権・特許権の侵害を発生するための分析に必要な、他人の著作物の複製		JEITA	
	20	著作隣接権者が侵害の探知を目的としてインターネット上をクロールし、自己が権利を有しないものも含め、あらゆる楽曲ファイルや動画ファイルをサーバーに複製する行為	インターネットサービス事業者	
G		著作物の対外的利用(自己の著作物に、他人の著作物が偶発的に写り込むこと等)		
	30	写真、映像や録音等の利用者が撮影等した投稿コンテンツに著作物が従属的に写りに込む場合など、偶然軽微な利用	延長問題フォーラム、MOF	
	31	家具や電化製品等のカタログとして、部屋での使用イメージを写真撮影した際、たまたま室内に飾られていた絵画等の著作物が写り込むこと	知財協	
		新商品とともに小さく著作物が写り込んだ写真を商品パンフレットとして複製する行為	インターネットサービス事業者	
	32	野外でテレビのニュース番組を収録していた際、現場で偶然音楽が微妙に漏れ、当該ニュース番組で放送されること	知財協	
	33	テーマパークで着ぐるみキャラクターと一緒に記念撮影した写真をインターネット上で公開すること	知財協、インターネットサービス事業者	
	34	キャラクターのイラストが描かれたTシャツを撮影し、雑誌に掲載すること	知財協	
	35	ぬいぐるみを抱いた赤ちゃんを撮影し、インターネット上で公開する行為	インターネットサービス事業者	
	36	機体にアニメキャラクターがあらわれた飛行機を撮影し、インターネット上で公開する行為	インターネットサービス事業者	
	37	屋外の看板やポスター、絵画が映り込んだ写真をインターネット上で公開する行為	インターネットサービス事業者	
	38	子供が描いたアニメの主人公の絵をインターネット上で公開する行為	インターネットサービス事業者	
	39	お祭りでお面を被った子供を撮影しインターネットに公開する行為	インターネットサービス事業者	
	40	レコード音源のBGMが入り込んだ動画をインターネット上で公開する行為	インターネットサービス事業者	
H		新たな技術、機器の研究開発・設計・製造・販売・故障原因分析の過程において、技術・機器の評価・検証に用いるための複製、上映、送信などの利用		
	41	データ圧縮・伸張技術の開発のために音楽CDを音源にすること	JEITA	
	42	AV機器、通信機器等の開発や性能試験のために音楽や映像を録音・録画したり、公衆送信したりすること	JEITA、知財協、インターネットサービス事業者	
	43	新聞用OCRソフトの開発にあたり、各新聞で使用されているフォントに対するOCRソフトの認識精度を高める目的で、新聞記事をスキャンすること	JEITA、知財協	
	44	製品開発中の翻訳ソフトの能力を確認する目的で、社内の研究室において専門書(例: 医学書)を複製し、翻訳させること	知財協、インターネットサービス事業者	
	45	音声認識における言語モデル、音響モデル等を作成するために、音声データをアーカイブする行為	インターネットサービス事業者	
	46	多数の著作物から、当該著作物を構成する要素である、言語、音等を抽出するために、データの収集・蓄積は行なうが、解析にまでは至らない行為	インターネットサービス事業者	
	47	音声認識アルゴリズムの精度検証のために、録画したテレビ番組の音声を用いる行為	インターネットサービス事業者)・プログラムの著作物の利用	
	48	プログラムの研究、性能の検証、障害発生時の原因追索を目的として行う当該プログラムの必要な限度の複製・翻案	JEITA	
	49	第47条の2に基づきプログラムの複製および翻案を第三者に委託して行わせること	知財協	
	50	旧サーバーから新サーバーにシステムを移行する際、新サーバーの信頼性が真極められるまでの間、旧サーバーにインストールされたソフトウェアを削除せず、新サーバーにインストールして使用すること	知財協	
	51	障害や脆弱性の発見、権利侵害の発見、既存プログラムの互換性の確認等の目的でリバース・エンジニアリングを行うこと	知財協	

区分	番号	内容	発言者
I	ネットワーキングサービスに関連する複製等		
	52	サーバーから端末上のRSSリーダーにデータをダウンロードする行為 (MCF)、ウェブ上のサイトのRSS情報を一定時間ごとに自動的に収集し、収集したRSS情報を利用者がいつでも見られるように、サーバー上に送信可能な状態におく行為	インターネットサービス事業者
	53	「Twitter」における「Retweet」のようにミニブログで他人の発信した情報を転送する行為	MCF, MIAU
	54	私的利用として作成したブックマークを共有する場合に著作物を複製する行為	MCF
	55	サーバー上の私的領域 (認証により他者の利用は不可) に著作物を保存して個人所有の複数の機器で利用する行為 (ネットワークサーバー上へのバックアップなどのクラウドサービス、タイムシフトやプレイシフトを目的として行われるサーバーへの複製、サーバーを介した不特定多数を対象としないファイル共有サービス、個人向けストリーミングサービス、バーチャルオフィス) 及びそのサービスの提供	MCF, MIAU, インターネットサービス事業者
	56	Internet Archive 日本版、ニュースや日記・ブログなどの表示形式変更・要約	MIAU
	57	タイムジフト、ブレイクシフト (ロクタク、MYUTA、まねきTV、録画ネットなどのサービス、不特定多数公開を前提としない録画代行サービス、リッピングなどのデータ形式変換サービス、P2Pに保存した音楽を携帯電話でも聞けるようにするアプリケーション)	MIAU
	58	Web全体もしくはその一部を収集 (クロール・リング)、保存して、アーカイブするサービスの提供	インターネットサービス事業者
	59	利用者が手動でURLを指定し、サーバー上に保存できるアーカイブサービスの提供	インターネットサービス事業者
	60	検索エンジン付随サービス (例: 検索キーワードと同じ文脈で扱われる語や類義語を抽出し、その語のみを表示するサービスの) の提供にあたり、ウェブ上の情報をクロールして収集する行為	インターネットサービス事業者
	61	インターネット上で商品 (美術品、写真の著作物) にあたらないもの) を販売する際に商品画像を掲載する行為	インターネットサービス事業者
	62	PC用に調製されたサイト画面を携帯端末でも表示できるように保存形式を変換して送信する行為	インターネットサービス事業者
	63	携帯電話用コンテンツをキャリアごとの条件に合うように変換をするために行なう複製	インターネットサービス事業者
	64	アクセス元のIPアドレス等を隠蔽保護するため中間サーバー経由でウェブアクセスを仲介、代行するサービスの提供	インターネットサービス事業者
	65	ニュースサイトからリンクを設定した先のサイトに大量のアクセスが集中した場合、サイトがダウンする可能性があるもので、あらかじめ当該サイトをキャッシュで保存し、ダウンした場合にはキャッシュ画面を代替表示させる行為	インターネットサービス事業者
J	個人の情報発信に伴う利用		
	66	権利者への経済的損害の極めてわずかな二次創作	延長問題フォーラム) [再掲]
	67	既存のコンテンツ、映像や音楽に対して、自分の工夫を加え、できたものを公衆に見せる、いわゆる「マッシュアップ」を行う行為	利用促進協議会, MCF, MIAU
	68	ブログで書籍やCD、DVDを紹介する際に書影やジャケットを掲載する行為	MIAU
	69	音楽を紹介する目的で試聴用音楽ファイルを掲載する行為	MIAU
	70	テレビ番組の批評や感想のため動画の一場面を静止画化した画像を掲載する行為	MIAU
	71	「Twitter」における「Retweet」のようにミニブログで他人の発信した情報を転送する行為	MCF, MIAU) [再掲]
	72	個人によるインターネットサービスを通じた転送行為	MIAU
	73	パロディ	MIAU
	74	個人が調査・研究目的で行う複製行為	MIAU
	75	ブログの著者から「自分のブログを印刷・製本してくれ」と依頼を受けた印刷業者が、ブログに寄せられた第三者のコメントも含めて印刷製本する行為	インターネットサービス事業者
	76	映画の感想とともにパンフレットの表紙、チラシ、ポスター等をインターネット上に載せる行為	インターネットサービス事業者
	77	自分の読んだ本の感想とともに、その本の表紙をインターネット上で公開する行為	インターネットサービス事業者
	78	歌碑や句碑を撮影し、インターネット上で公開する行為	インターネットサービス事業者
	79	新聞・雑誌ではなく、インターネット上で公表された時事問題に関する論説を転載する行為	インターネットサービス事業者
	80	屋外に設置されている街路地図を撮影し、インターネット上で公開する行為	インターネットサービス事業者
K	81	シンボリズム等でのハネラーの発言を要約筆記してインターネット上で公開する行為	インターネットサービス事業者
	行政機関による著作物の利用		

区分	番号	内容	発言者
	82	意匠審査における新規性判断のために、特許庁が雑誌、カタログ、ホームページに掲載されている意匠公知資料を特許庁がインターネット上で公開(公衆送信)すること	知財協、インターネットサービス事業者
L	その他		
	83	着メロ作成時[MIDIファイルや音源ファイル]を複製又は利用するなど、コンテンツ制作工程において行う一時的な複製等	84～100 インターネットサービス事業者 MCF
	84	保育園の教諭が園児に絵本を読み聞かせる行為	
	85	保育園の教諭が絵本を元に紙芝居を作り、園児に見せる行為	
	86	保育園の教諭がピアノを演奏し、園児と歌う行為	
	87	大学が国語の入試問題を作成する際、作問候補の小説を数本複写した上で、どれを問題に採用すべきか会議で検討する行為	
	88	ピアノ教師が生徒の前で模範演奏する行為	
	89	美大の教官が講義で使用した写真資料をeラーニングでも配信する行為	
	90	大学受験予備校の講師が受験の口占あわせで替え歌を作る行為	
	91	英会話教師が教室で洋楽を再生し、生徒に聴き取らせる行為	
	92	自宅で録音したピアノ演奏を、オーディションの審査員たちに聞いてもらう行為	
	93	アニメキャラクターの塗り絵大会が催され、上手な塗り絵を商店街に掲示する行為	
	94	人気アニメのキャラクターの雪像を作り、校門の横で展示する行為	
	95	図書館利用者が複写箇所を図書館に申告することなく、館内のコピー機で複写する行為	
	96	小児科の看護士がキャラクターの人形を手縫いで作り、入院児童にあげる行為	
	97	友人から借りた複製機を用いてCDを自分の個人鑑賞用に複製する行為	
	98	合唱コンクールで生徒が歌っている様子を教師がビデオカメラで撮影し、コンクールに参加した生徒の求めに応じて、録画ビデオをダビングしてあげる行為	
	99	住民代表が情報公開制度に基づいて入手した図面資料を地域住民にFAXで送信する行為	
	100	新聞社が画家の死亡記事に添えて、その画家の有名絵画作品を紙面に掲載する行為	

CODATA部会報告

## 平成21年度科学技術データベース懇談会報告

馬場哲也<sup>1\*</sup>、山下雄一郎<sup>1</sup>、岩田修一<sup>2</sup>

Tetsuya BABA<sup>1\*</sup>, Yuichiro YAMASHITA<sup>1</sup>, Shuichi IWATA<sup>2</sup>

1 産業技術総合研究所

National Institute of Advanced Industrial Science and Technology

2 東京大学

The University of Tokyo

\*連絡先著者 Corresponding Author

最近是一部の科学技術データ、例えば大気中のCO<sub>2</sub>含有率や自然環境の平均温度上昇率などが、経済活動や国際交渉に極めて大きな影響を及ぼすようになった。それにもかかわらず、科学技術データベース(以下DB)の信頼性やDBの計画性ある整備、人材育成への社会や政府の関心が低いといわざるを得ない。

そうした背景を鑑み、科学技術DBに直接関わる方々で自由闊達な懇談を行って、現状の問題を知ると同時に将来への働きかけの可能性について、意見交換の場として、CODATA部会はH21年度より「科学技術データベース懇談会」を発足させた。発足時の委員は化学・材料分野のデータおよびデータベース研究に携わる研究者によって構成された。委員は地質、ライフサイエンス、環境等幅広い分野の研究者を順次対象にする予定である。学会員の皆さまに懇談会の討議内容を広く周知することで、科学データおよび科学技術データベースに関心を抱いて頂けるよう、H21年度に開催した2度の懇談会の概要を報告する。

第1回懇談会は2009年11月18日に東京ステーションカンファレンスにて開催された。第1回の冒頭において、懇談会の目的は、①関係者間の問題点の共通理解と連帯感の形成、②将

来のDB像とそのアウトカム評価の在り方、③DBの中核となる機関の形成にあるとした。次に、各参加委員からは担当研究分野についての概要報告があった。現状を改善するための取り組みについていくつか提案された項目のみを以下に記す。

- ・データベースのイメージの変換
- ・新しいデータベースへのチャレンジの必要性
- ・データベース会社(仮)の可能性
- ・The Forth Paradigm

これらの提案を基に、データベースの成功事例とその要因、データベースの連携可能性、科学技術データを所管する機関の形成について討議が行われ、討議内容について参加委員の共通意識を形成するに至った。

第2回懇談会は2010年1月22日に東京大学本郷キャンパスにて開催された。第2回には、地質分野の研究者を迎え入れ、議論の幅を広げた。地質分野では、DBを明確なアウトプットとして位置付けているとともに、研究者自身の意識もデータベースを向いている点、全世界的な地質情報の動向としてデータベースの連携が進んでいる点について、データベースの成功モデルとして位置づけられた。

第2回懇談会では、「データベースの可能性」

について討議を行った。討議の中では既に整備されたデータベースを有機的に連携し、専門家でなくともデータを利用できるデータベースの構築が必要であるとされ、「社会・経済ダイナミックデータベース」、「環境に関してのデータベースの複合」、「医工連携のための複合データベース」が例示された。また、学術研究の対象として、データベースの広域利用(分野相互間)のためのさまざまなDBの整備と利用の一貫した規格化、標準化についての研究を行ったり、社会的関心の喚起を行ったりする必要があるとの認識で一致した。

平成21年度の懇談会を通じて、社会的関心の喚起のため、Webサイトを立ち上げデータおよびデータベースの情報を発信する取組の実施と、各委員が担当するデータベースを中心にDB間の連携を図り、冒頭に記した社会問題に対して的確な情報を提供するサービスの構築を検討することで合意がなされた。

本活動は平成22年度にも継続する予定であり、第3回懇談会は4月に東京において開催される。



## 第7回(2010)論文賞の選定について

第7回(2010)の論文賞の選定を行います。第5回(2008)論文賞の選考方式に基づき、学会員が直接投票で選びます。論文賞推薦委員会は昨年度からの継続です。委員は、安永副会長(委員長)、根岸会長、国沢常務理事、長塚常務理事の4名です。

### 選定の日程

- 1 論文賞の候補の推薦. 本学会員(正会員, 賛助会員)は, 推薦委員会に対して論文賞にふさわしいと思われる論文をその理由をつけて推薦する.
  - ・推薦開始 :2010年2月25日
  - ・推薦締切り:2010年3月20日
- 2 推薦委員会は, 会員からの推薦論文が多数の場合は一次選考を行い, また少数の場合は推薦委員会により追加推薦を行って, 候補論文を決定する.
  - ・候補決定 :3月25日
- 3 これら論文賞候補論文とその推薦理由を学会ホームページおよびメルマガ等に掲載し, 会員に投票を依頼する. なお, 推薦者の名前, 人数などは公表しない.
  - ・投票開始 :3月25日
  - ・投票締切り:4月23日
- 4 投票の結果, 最多得票の論文を論文賞授賞論文とする. ただし, 推薦委員会は得票数や論文内容などを勘案し, 得票数第2位の論文についても論文賞とすることができ
- 5 選定結果発表
  - ・授賞式: 次期総会(5月15日(土))において

### 推薦対象論文(下記リストは, 掲載順)

- 1 吉廣卓哉, 井上悦子, 田部浩子, 中川優:生活習慣病予防教室における効果傾向抽出のためのデータマイニングシステム, Vol. 19, No.1
- 2 森山賀文, 小野智司, 中山茂:免疫アルゴリズムを用いた複数画像探索と書籍特定への応用, Vol. 19, No. 3
- 3 高田良宏, 笠原禎也, 西澤滋人, 森雅秀, 内島秀樹:非文献コンテンツのための可視性と保守性に優れた学術情報リポジトリの構築, Vol. 19, No. 3
- 4 山地一禎, 片岡俊幸, 宮地直人, 曾根原登:Office Open XMLに対する長期署名の付与, 2009.12.22(オンライン公開)

<注> これらの論文は学会誌の他, オンライン(J-Stage)でも論文全文を参照できる.

<http://www.jstage.jst.go.jp/browse/jsik/-char/ja/>

#### 推薦方法・締め切り

推薦する論文について、400～800字程度の推薦理由を付して、2010年3月20日までに、学会事務局(jsik@nifty.com)、および推薦委員会(yasunaga@nihu.jp)あて電子メールで送信する。

形式自由。ただし、SUBJECT欄に「論文賞候補推薦状」と明示すること。

#### 意見募集

昨年度も同じでしたが、単年度内では論文賞の対象とする論文が少ないので、年度以降の論文賞のあり方を検討すべきと言う意見が、常務理事会で提案されています。

そこで、論文賞対象とする論文の範囲などについて、会員の意見を求めます。

例えば、2年置きに選定、あるいは複数年度に渡って(前年度の論文も重複して)対象にするとかの案も出ています。

ご意見は、事務局(jsik@nifty.com)まで。

## 情報知識学会 第 18 回 (2010 年度) 年次大会 発表論文募集について

2010 年 1 月 15 日

第 18 回 (2010 年度) 年次大会実行委員会

実行委員長 梶川裕矢 (東京大学)

実行委員 高久雅生 (物質・材料研究機構)

江草由佳 (国立教育政策研究所)

森純一郎 (東京大学)

情報知識学会では下記の期日・場所で、総会と共に研究報告会を開催いたします。

- 日程: 2010 年 5 月 15-16 日 (土・日)
- 会場: 東京大学本郷キャンパス

発表論文を下記要領で募集いたしますので、会員の皆様には、奮ってご応募頂きますようお願い申し上げます。

### 1. 募集分野

1. 情報知識の構造解析、モデル化、可視化、知識発見
2. 情報・知識の表現、生産、組織化、検索、提供
3. 電子出版、電子図書館
4. マルチメディア、電子ミュージアム
5. 用語、シソーラス
6. 知識情報の流通と知的所有権
7. 専門分野における情報の品質管理、基準化
8. インターネット、セマンティックウェブ、Web2.0 など
9. その他情報知識学に関連する諸研究・開発

### 2. 応募方法

- 登録方法: 下記のサイトから必要情報を登録。

<http://tinyurl.com/yg5o6ek>

- 応募期限 : 2010 年 2 月 28 日 (日)
- 採択可否通知 : 2010 年 3 月 5 日 (金)
- 原稿提出期限 : 2010 年 4 月 2 日 (金)

### 3. 論文執筆・発表について

1. 原稿の体裁は、学会誌の論文執筆要領に準拠してください。
2. 発表時間は質疑応答を含めて 30 分です。論文提出がないと発表できません。
3. 登壇発表者は当学会員に限ります。当日入会も可能です。
4. 本論文も WEB 上での提出を予定しています。具体的な方法は採否決定時に改めて連絡いたします。

### 5. お問い合わせ先

- 〒113-8656 東京都文京区弥生2-11-16 TEL/Fax: 03-5841-1161 E-mail: jsik2010(at)ml.nims.go.jp

(会告)

## 会員増加のための諸制度改定のお知らせ

### 入会金廃止、年度途中入会時の会費割引、若手・シニア向き会員種別の新設

会長 根岸正光

2009年5月の本学会総会において、新規会員の入会促進と現会員の退会抑止を目指した対策の策定が理事会に付託されました。常務理事会を中心にその原案の検討を進め、2010年1月15日の持ち回り理事会において、会員種別と会費に関して次のような改定を行うことが決定され、即日実施されました。改定の趣旨と内容は以下にのべるとおりです（別表参照）。会員各位におかれては、この趣旨をご理解の上、会員増強について格段のご協力を賜りたくお願い申し上げます。

#### 1 入会金の廃止

これまで新規入会者からは入会金をいただいております。もっともホームページを経由した入会申し込みの場合、その特典として入会金免除としておりましたし、年次大会や情報知識学フォーラム参加者についても同様の特典措置で入会案内を行ってきたところがあります。そこでこの際、新規入会を一層促進するため、この入会金自体を廃止して、年会費のみで入会できることに致します。

#### 2 年度途中入会時の会費割引の導入

これまで、年度途中で入会の場合にも、1年分全額の会費を納めてもらうことになっていましたが、これを四半期単位の割りとして、例えば年度の第4四半期での入会では4分の1でよいというように、割り引くことにします。入会金廃止と併せて、新規入会の障壁を少しでも低くしようとするものです。

#### 3 「シニア会員」種別の新設

本学会は今年創立23年目を迎え、創立当初からの会員も多くおられます。これらの先輩会員の方々にあっては定年退職などもあり、退職を機に退会を考えられる方々もおられます。本学会としては、こうした諸先輩にも、引き続き正会員として活動していただきたいところではありますが、それでは何かと荷が重いと考えられる方について、とりあえず年

会費を半減して、シニア・メンバーとして本学会にとどまっていたいただき、引き続きご助言、ご指導を賜るようにするため「シニア会員」の種別を設けました。これには特段年齢要件はないので、事情により退職された会員はどなたでもこの種別を選択できます。なお、役員の被選挙権はそのまま継続され、役員に当選し、これを受諾された場合には正会員に復帰していただくことにします。また、復職された場合にも正会員に戻っていただくようお願いいたします。

#### 4 「ユース会員」種別の新設

年次大会での発表を機会に学生会員として加入された大学院生等の方々と、卒業後に大学、研究所等に就職された方は、引き続き正会員として参加いただく場合が多いのですが、企業などに就職して必ずしも研究に直接携わらなくなるような場合には、正会員として継続参加されずに退会を考える方も多いようです。本学会は情報知識に関わる諸問題を幅広く扱い議論する場であり、研究者の視点だけに止まらず、多様な業務に関わる実務者の観点からの議論も重要であります。従って、研究職でない方々にも参加いただくのは大いに有意義であるので、上のような事情にある35歳未満の方々に対して、より参加しやすくするため、会費を半減し学生会員と同額とする「ユース会員」の種別を設けました。

また、卒業後に研究職に就いた場合であっても、若手の場合、ポスドク等、正規雇用職員でないことも多いかと考えられます。「ユース会員」には、こうした方々の経済的負担を多少なりとも軽減して、若手研究者の参加を促すという趣旨もあります。

この種別は35歳未満という年齢要件だけですので、学生会員を経ずに新規加入の場合でも選択できます。なお、シニア会員と同様、役員の被選挙権もあり、役員に当選し、これを受諾した場合には、正会員になっていただくことにします。

#### 5 役員の選挙権、被選挙権、総会議決権の拡大

役員の選挙権、被選挙権および総会の議決権については、従来すべて正会員に限るものとして運用してきましたが、これを次のように他の会員種別にも拡大して適用します（別表参照）。

役員の選挙権は正会員の他、協賛会員、特別協賛会員にも各1票あて割り当てます。役員の被選挙権は新設のシニア会員、ユース会員に加えて学生会員にも拡大し、若手の学会運営への積極的参加を期待します。なお、学生会員が役員に当選、受諾の場合は、学生会員のままで役員を務めていただくことにします。

## 6 論文投稿の促進

論文投稿はすべての会員種別（協賛会員、特別協賛会員にあっては当該団体に所属する者）について可能ですので、積極的投稿をお願い致します。なお、論文投稿を会員に限らずに、ひろく一般からも受け付けるような例が特に大規模学会ではひろがりつつある状況です。こうした方向について、本学会としてどのように考えるべきか、難しい問題で、編集委員会を中心に議論してもらっていますが、会員各位においてもご意見をお寄せ下さい。

ところで、上記のような措置は当面導入可能なものとして、以前から総会等の場で議論されてきたものであり、会員各位におかれては、この趣旨を踏まえて、会員増強と論文投稿促進に心がけていただくようお願い致します。また、今回の改定に止まることなく、さらに積極的、効果的な会員増強策、学会活動の活性化策について、身近な周辺の状況を勘案しつつご検討いただき、各種の提案をお寄せ下さるようお願い致します。

（別表） 会員種別ごとの会費、選挙権等の一覧（2010 年 1 月 15 日改定実施）

会員種別	論文投稿	役員 選挙権	役員 被選挙権	総会 議決権	年会費 (円)	条 件
正会員	○	○	○	○	8,000	
学生会員	○	×	○	×	4,000	
ユース会員	○	×	○	×	4,000	35 歳未満は本人希望により選択可。 35 歳からは正会員。
シニア会員	○	×	○	×	4,000	退職者は、本人希望により選択可。 復職時は正会員。
名誉会員	○	×	×	×	0	
協賛会員	○	○	×	○	1 口 30,000 以上	当該団体所属の個人は論文投稿可。
特別協賛会員	○	○	×	○	5 口 150,000 以上	当該団体所属の個人は論文投稿可。

「情報知識学会誌」投稿規定

2002年8月27日 制 定

2003年3月19日 一部改定

2006年8月 1日 一部改定

0. 情報知識学会誌編集規程による本会機関誌「情報知識学会誌（以下、会誌という）」への投稿に関する事項は、この規定の定めるところによる。

1. 投稿資格

投稿者の少なくとも1人は本会員でなければならない。ただし、編集委員会による依頼原稿の場合にはこの限りではない。

2. 投稿原稿

2.1 広い意味での情報知識学に関連し、またその発展に貢献するもの（情報／知識の収集、整理、蓄積、検索および各種解析、利用などに関するもの）とする。刊行時において未発表の原著でなければならない。本会誌の記事の種類を以下に示す。

2.2 投稿者は会誌記事の種類を明記して投稿しなければならない。ただし、編集委員会に変更することがある。

(1) 研究論文 (Research Paper) : オリジナルな研究論文で、内容の主要な部分が学術論文として他に公表されていないもの。

(2) 事例／調査報告 (Report) : 情報知識学に関連したシステムなどの開発、利用、調査に関するもの。資料も含む。

(3) 解説／展望 (Review) : 情報知識に関連した特定分野の論文や学説などを総括、解説、紹介、あるいは技術動向などを展望したもの。技術、研究上の処理、解析方法などに関する解説。

(4) 論談 (Proposal Paper) : 情報知識学に関連した新たな意見の表明、提案など。

(5) 討論 (Discussion) : 本会誌に掲載された論文についての学術的な討論。

(6) 研究速報 (Notes) : 技術、手法、新事実などの簡単な報告。

(7) 講座 (Lecture) : 情報知識学の各分野に関する基礎理論、技術の適用などについて、テーマを定めて系統的に説明するもの。

(8) 学会記事 (News) : 本会の事業、運営などの報告、記事、資料など。

(9) ニュース、お知らせ (News) : ニュース、お知らせ。最近刊行された単行本やモノグラフの紹介

(10) 講演 (Lecture) : 特別号などにおける講演資料。

(11) その他：編集委員会が適当と判断したもの。

2.3 会誌記事の種類のうち、(1)から(6)までは査読を行う。その他については編集委員会で編集を行う。

3. 投稿原稿

3.1 原稿の形式

(1) 最初の投稿時

原則として、以下の体裁で作成された電子ファイル (PDF形式) を電子メールに添付し

た投稿とする。体裁は、刷り上り原稿を想定したレイアウト（A4判，2段組，20字×46行×2段）にして、図，表は希望の位置に配置すること。その他の執筆に関する詳細は「執筆要領」を参照のこと。

## (2) 採択決定後の原稿

PDFとその元になったファイル（Word fileなどで編集可能なもの）。

## 3.2 原稿の制限

### (1) 原稿の長さを原則として次のように制限する。

研究論文，事例／調査報告，解説／展望，論談：刷り上がり20ページ以内

討論，研究速報，講座：刷り上がり6ページ以内

ニュース他：刷り上がり2ページ以内

### (2) 冊子体の図原稿（原図）の大きさはA3判を越えないものとする。

### (3) 原則として，図版も含めてモノクロ印刷とする。ただし，カラーでなければならない図版を使用する場合は，別途編集委員会と相談する。なお，カラーページやページを超過する分については，印刷費を著者の全額負担とする。本学会誌はJ-STAGEから電子ジャーナルとしても公開するので，カラーの図をWeb上の電子付録とすることができる。また，冊子体よりもより詳細な図表やさらには動画も電子付録とすることが可能である。電子付録はすべて無料で利用できる。

### (4) 使用言語は日本語または英語とする。

## 4. 原稿の採否

投稿原稿の採否は，専門家による査読の後，編集委員会において決定する。

## 5. 査読のプロセス

学会員の中から編集委員会が指名した査読者2名によって査読を行う。内容によっては，編集委員会は著者に照会し，原稿の修正を求めたうえで，再査読を行うことがある。

## 6. 校正のプロセス

採択が決定した投稿原稿は，掲載原稿として著者に校正を依頼する。著者による校正は原則として1回とする。その際，字句の修正以外は原則として認めない。

## 7. 別刷

別刷（抜刷）は著者の実費負担とする。希望部数を事務局に申し出ること。

## 8. 投稿の手続き

最初の原稿投稿時には下記のファイルを電子メールに添付する。

### 8.1 必要ファイル

a. 投稿原稿整理カード：ホームページからコピーして，必要事項を記入したテキストファイル。

b. 論文原稿のPDF形式ファイル（図，表を含む）

### 8.2 原稿の送付先

学会誌編集委員会委員長 E-mail: kunisawa@rs.noda.tus.ac.jp

なお，以下の2つのアドレスにもCCメールとして送ること。

学会誌編集委員会副委員長 E-mail: ashino@acm.org

情報知識学会事務局 E-mail: jsik@nifty.com

### 8.3 原稿の受付

事務局が原稿を受け取った日を受付日とする。受付の確認を1週間以内に投稿者の連絡先にE-mailで通知する。不備のある投稿原稿は返送し，再提出するものとする。

### 8.4 著者は査読候補者リスト（5名程度の住所、所属、電子メールアドレスを記入した



もの)を提出できるものとする。

9. 原稿提出期日

投稿は随時とする。ただし、特集号などは除く。

10. 著作権

10.1 機関誌『情報知識学会誌』に掲載された論文(電子版を含む)の著作権(著作財産権, copyright)は情報知識学会に帰属する。

10.2 掲載論文は冊子による出版の他、電子的に蓄積し、本会が行う情報提供サービスなどを通じて公開する。

10.3 本学会誌に掲載された執筆内容が第三者の著作権を侵害するなどの指摘がなされた場合には、執筆者がその責任を負う。

11. 規定の改訂

11.1 本規定の改訂は、編集委員会の議を経て、理事会の承認を得なければならない。

12. 施行

12.1 本規定は2006年7月1日より施行する。

12.2 本規定の施行により、現行規定(第5版(暫定板)2003年3月)は廃止する。

13. 改訂履歴

2003年3月19日一部改訂。「10. 著作権」に、10.3項を追加。

2006年8月 1日一部改訂。投稿手段を郵送から電子メールに変更。

## 「情報知識学会誌」執筆要領(中央揃え, MSPゴチ, 16pt)

### Title in English (centered, Times New Roman, Bold, 16pt)

国沢隆<sup>1\*</sup>, 芦野俊宏<sup>2</sup> (中央揃え, MSPゴシック, 12pt, 連絡先の著者に'\*'をつける)  
Takashi KUNISAWA<sup>1\*</sup>, Toshihiro ASHINO<sup>2</sup> (centered, Times New Roman, 12pt)

1 東京理科大学(左揃え, MS明朝, 10pt)

Tokyo University of Science(left-aligned, Times New Roman, 10pt)

〒162-8601 東京都新宿区神楽坂1-3(左揃え, MSP明朝, 10pt)

E-mail: kunisawa@rs.noda.tsu.ac.jp(left-aligned, Times New Roman, 10pt) [アドレスの記入は任意]

2 東洋大学

Transdisciplinary Research Integration Center, Toyo University

〒112-8606 東京都文京区白山5-28-20

E-mail: ashino@acm.org

\*連絡先著者 Corresponding Author

この部分には要旨を記述する。研究論文, 事例／調査報告, 解説／展望, 論談の原稿には, 和文および英文で要旨をつける。和文要旨の長さは400 字以内とする。要旨中には, 図, 表, 数式などを用いない。本文中の図, 表, 数式, 文献などを番号で引用しない。情報知識学会誌に投稿する場合は, この執筆要領に従って作成すること。ただし, フォントについては, 明朝・ゴシックを基本とする。この文書では特定の環境に依存した指定があるが (MSP明朝やMSPゴシック等), 単に明朝・ゴシックと指定するとウエイトの違いなどのかかなり体裁の異なったものができてしまう可能性をできるだけ回避するために指定しているものであるため, あくまでも目安とすること。(MSP明朝 11pt.)

English abstract. within 200 words. If you submit a Research Paper, Research Note,....., you have to add a abstract. In the abstract, you should not cite figures, tables and formula expression in the article. (Times New Roman. 11pt).

キーワード: キーワード1, キーワード2, キーワード3, キーワード4, キーワード5,(MSP明朝,10pt)

Keyword1, Keyword2, Keyword3, Keyword4, Keyword5(Times New Roman、10pt)

(研究論文, 事例／調査報告, 解説／展望, 論談, 討論, 研究速報, 講座にはキーワードをつける。和文および英文でそれぞれ5 個程度, 和文と英文のキーワードは, 対応することが望ましい。キーワードはカンマ(,)で区切る。)

## 1 一般的な事項 (MSPゴシックまたはTimes New Roman 14pt, Bold)

本文は二段組、MS明朝またはTimes New Roman、11ptを用い、ページ設定は、用紙をA4縦として、余白は上35mm, 下25mm, 左右25mmとする。

本会誌への投稿は、「投稿規定」に従い、投稿原稿は本執筆要領に従って作成されなければならない。本会誌の投稿原稿の種類には、研究論文、事例／調査報告、解説／展望、論談、討論、研究速報、講座、本会記事、講演、ニュース、その他がある。

## 2 日本語原稿の構成

### 2.1 全体構成 (MSPゴシックまたはTimes New Roman, 12pt, Bold)

- ・ 標題 (和文および英文)
- ・ 著者名 (和文およびローマ字, ローマ字による著者名は、名、姓の順で、姓は全て大文字を使用する。)
- ・ 所属 (和文および英文による所属機関名)
- ・ 住所 (和文による所属機関の住所, E-mail。)
- ・ 本文 (和文または英文)
- ・ 文献、付録など (和文または英文)
- ・ その他 (とくに長い論文の場合、読者の便宜を考えて内容目次を付してもよい。ただし、章、節の見出し程度とする。)

### 2.2 本文 (Body)

#### 2.2.1 構成 (MSPゴシック, 12pt, Bold)

章、節などの構成は、第1 レベルは1, 2, …, 第2 レベルは1.1, 1.2, …, 第3 レベルは1.1.1, 1.1.2, … のようにする。

#### 2.2.2 脚注

脚注はできるだけ避ける。止む無く使用する場合は簡潔な文とする。

#### 2.2.3 図および表

図、表にはそれぞれ通し番号をつける。図1 (Fig. 1), 図2 (Fig. 2), … 表1 (Table 1), 表2 (Table 2), … など。

#### 2.2.4 数式, 化学式

a. 数式 (独立式), 化学式は、段落外で記

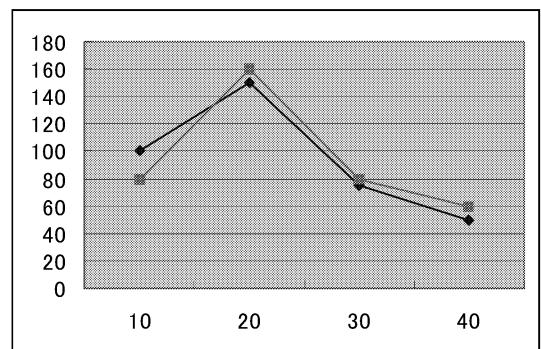


図 1 通し番号とともに説明文 (キャプション) をつける。キャプションの位置は、図は下部に、表は上部とする (図表番号は MS 明朝または Times New Roman, 10.5pt, Bold, キャプションは MS 明朝または Times New Roman, 10pt)

述されているものも本文中で一回は参照する。

b. 数式には、通し番号を振る。

#### 2.2.5 リスト (または箇条書き)

- a. 記号なしリスト。
- b. 記号つきリスト。リストの記号は、数字、アルファベット、記号を用いることができる。ただし、これらの混在した使

表1 一段組みにした図表のキャプションは中央に揃える

Fertilizer	1977	1991	1992
Mineral	1000	204	135
Nitrogen		87	65

用は避ける．アルファベットは1 論文中では大文字，小文字の使い分けをしない．

c. 複雑化を避け，せいぜい2段（親子関係）のリストとし，ネストを跨ぐ順序づけを用いない．

## 2.2.6 注記および参考文献

本文中で少なくとも一回は参照すること．次記のように，[通し番号]として参照し，タイトルなどでの参照は避ける[1]．

## 2.3 後付け(End)

### 2.3.1 注記および参考文献

a. 注記または参考文献には，参照順に通し番号を付し，本文の最後に番号順にまとめて記述する．章番号は用いない．章題は「参考文献」とする．

b. 1つの番号には1つの注記または参考文献を対応させる．

c. 注記中には参考文献を含めない．注記はできる限り簡潔に表現すること．

d. 参考文献の記述形式は，以下の形式を満たさなければならない．

e. URLを参照してもよいが，移動または削除される可能性があるので，極力避ける．原著がURLでのみしか参照できない場合など，やむをえない場合は用いてい．その場合，参照時点でのハードコピーを保管しておくなど，参考文献へのアクセス手段を確保するよう努力しなければならない．

#### 【参考文献の形式】

#### 1. 雑誌中の1論文

[引用通し番号] 著者名：論文名，雑誌名，巻号，掲載ページ，出版年，その他．

#### 2. 図書1冊

[引用通し番号] 著者名：書名，版表示，出版地，出版社，総ページ数，出版年，その他．

#### 3. 図書の1部

[引用通し番号] 著者名：論文名，書名，版表示，出版地，出版社，掲載ページ，出版年，その他．

#### 4. 会議報告

[引用通し番号] 著者名：論文名，書名（会議名），版表示，編集者名，会議開催地，会議開催年，会議開催機関，出版地，出版社，掲載ページ，出版年，その他．

#### 5. インターネット上の論文

[引用通し番号] 著者名や標題など可能な限り詳細な書誌事項，URL，参照年月日．（単なるホームページなどは参考文献にしないこと）．

#### 【参考文献の記述】

#### 1. 著者名，編集者名の記述

(1) 個人著者名は，姓，名の順に記述する．欧文著者名は，カンマ（,）で姓，名を区切る．

(2) 複数著者の場合は，各著者をセミコロン（;）で区切る．

(3) 翻訳図書などの翻訳者名の場合は，著者名の後に括弧（ ）に入れて記述する．

#### 2. 論文名，書名の記述

(1) 論文名、書名は、和文の場合はかぎ括弧（「」）、欧文の場合はダブルクォーティション（” ”）に入れて記述する。

(2) 図書中の一部を引用した場合の書名は、和書の場合は二重かぎ括弧（『』）に入れ、欧文の場合はイタリック体で記述する。

### 3. 掲載ページの記述

(1) 論文の場合は、開始ページと終了ページを記述する。「pp. 開始ページ- 終了ページ」とする。

(2) 図書の場合は、総ページ数とする。「総ページ数p.」とする。

## 3 文章と文体

1. 文体はひらがなと漢字による口語常態（である調）とし、現代かなづかいを用いる。
2. 漢字は当用漢字とする。ただし、固有名詞や学界で広く用いられている慣用の術語はこの限りではない。
3. 句読点その他には「，」「．」を用いる。
4. 本文中の人名には敬称をつけない。ただし、謝辞の人名はこの限りではない。
5. 数量を表す数字はアラビア数字とする。
6. 数式は印刷に便利のように十分注意して記号を記すこと。原則として数量（変化量）を表す記号はイタリックとする。
7. ローマ字の人名の姓は大文字体とする。
8. 固有名詞で読み誤るおそれのあるものにはふりがなをつける。
9. 英数字は原則として半角英数文字で記述する。

## 4 英文原稿

英文による投稿原稿の場合も、原則として

和文による投稿原稿の諸規定に従う。英語圏以外の著者の場合、著者名表記にその国語による表記を認めるが、可能な限り英文表記とする。

・研究論文、事例／調査報告、解説／展望、論談、討論、研究速報などの原稿は英文でもよい。

・英文原稿は語学的に難点の少ないものであることを必要とし、著者の責任において完全を期する。

・英文原稿には、英文による要旨 200 語程度、ならびに日本語による 400 字以内の要旨をつける。ただし、著者が日本語を理解できない場合は日本語要旨を省略できる。

## 5 J-STAGEの電子付録

本学会誌の記事のうち学術的なものはJ-STAGEからも公開する。したがって、カラーの図や冊子体よりも詳細な図表を電子付録としてWeb上で公開可能である。動画なども電子付録として受け付ける。ただし、査読論文の電子付録は査読の対象となり、掲載決定後に内容の変更はできない。

## 6 その他

原稿は和文または英文によるものとする。文章は語学的に難点の少ないものであることとし、著者の責任において完全を期する。編集委員会は語学的校正を行わない。

## 7 要領の改訂

本要領の改訂は、編集委員会の承認を得なければならない。

## 8 施行

本規定は2002年8月27日より施行する。

## 9 改訂履歴

2003年5月2日一部改訂．英語要旨の長さを500語から200語に変更．図，表のキャプション位置を訂正．

2009年7月24日一部改訂．表題と本文フォントおよび著者表記の変更．

## 謝辞

本文の最後に続けて記述する．章番号は用いない．章題は「謝辞」とする．最終原稿時に記述することが望ましい．

## 参考文献

[1] 藤原譲：「情報知識学試論」，情報知識学会，Vol.1, No.1, pp.3- 10, 1990.

[2] 原正一郎；安永尚志：「国文学研究支援のためのSGML/XML データシステム」，情報知識学会，Vol.11, No.4, pp.17- 35, 2002.

[3] Fujiwara, Shizuo: “East-West Communication and Information Transfer — Coordination of Specificity”, Journal of Japan Society of Information and Knowledge, Vol.4, No.2, pp.11-18, 1994.

[4] Ellis, David (細野公男監訳，斎藤泰則，鈴木志元，村上泰子訳)：「情報検索論」，丸善，180p.，1994.

[5] 根岸正光：「学術情報の流通と利用」，『情報学とは何か』情報学シリーズ3, 丸善，pp.43- 69, 2002.

[6] 名和小太郎：「デジタル図書館と著作権」，デジタル図書館，No.4, <http://www.dl.ulis.ac.jp/DLjournal/No4/nawa/nawa.html> (2002年8月27日参照)

## 事務局からのお知らせ

### 〔 1 〕 会員種別の変更手続きについて

従来は正会員・学生会員・名誉会員・協賛会員・特別協賛会員の五種でしたが、このたび新たに、ユース会員とシニア会員が新設されました。詳細は本号掲載の「（会告）会員増加のための諸制度改定のお知らせ」をご覧ください。正会員または学生会員からシニア会員またはユース会員へ変更を希望されるかたは、事務局へ電子メールその他でお申し出ください。様式は問いません。ご不明の点はご遠慮なく下記の事務局へお問い合わせ願います。

### 〔 2 〕 平成 22 年度年会費の納入をお願いします

平成 22 年度は、本年 4 月 1 日から来年 3 月 31 日までの 1 年間です。郵便局または銀行の下記口座へ 5 月末までにお振込ください。1 年間の年会費は正会員 8 千円、学生会員・ユース会員・シニア会員は 4 千円です。過去数年分未納のかたは合計額を納入してください。請求書が必要な場合、その旨を事務局へ電子メールその他でお知らせくだされば郵送いたします。

#### 1. 振込先（振込手数料はご本人負担をお願いします）

- a. 郵便振替口座 00150-8-706543 情報知識学会（代表 根岸正光）
- b. ゆうちょ銀行 O一九店（ゼロイチキュー店）当座 0706543 情報知識学会（代表 根岸正光）

#### 2. ご自分が納入した年月日の確認方法

情報知識学会から郵送された封筒の宛名ラベルをご覧ください。〔 〕 内に過去 4 年間、ご自分の納入日が印字されているので確認できます。納入年（西暦の下 2 桁）、月（2 桁）、日（2 桁）の 6 桁です。年会費を滞納している場合は、〔未納〕と表示してあります。金融機関へ振り込まれてから事務局へ通知が届き、宛名ラベルに印字、発送するまで 10 日ほどかかりますので、ご了承ください。

### 〔 3 〕 情報知識学会メールマガジンをご愛読ください

電子メールアドレスを事務局へ登録されているかたへは情報知識学会メールマガジンを毎月配信し、本学会の活動内容や関連情報を迅速にお伝えしています。以前に登録されたかたでも、情報知識学会事務局から最近の 1 ヶ月以内にメールを 1 通も受信していない場合は、不達が予想されますので、再度、アドレスを事務局へお知らせください。（jsik@nifty.com）

### 〔 4 〕 電話でのお問い合わせ

事務局の業務は土日祝日を除き、月曜から金曜日までの毎日行っています。お問い合わせなどの電話は、できるだけ午後 1 時半から 5 時までをお願いします。連絡には電子メールや F A X も、どうぞご利用ください。

入会ご希望のかたには入会申込書を、郵送または F A X 送信でお届けします。

## 情報知識学会事務局

〒110-8560 東京都台東区台東 1-5 凸版印刷棟内

TEL:03-3835-5692 FAX:03-3837-0368

E-mail:jsik@nifty.com URL:http://www.jsik.jp

---

---

## 情報知識学会誌 編集委員会

編集委員長 国沢 隆 東京理科大学  
副編集委員長 芦野 俊宏 東洋大学  
編集委員

相田 満	国文学研究資料館	石井 守	情報通信研究機構
石塚 英弘	筑波大学	岩田 覚	東京大学
内田 努	北海道大学	宇陀 則彦	筑波大学
江草 由佳	国立教育政策研究所	大久保 公策	国立遺伝学研究所
岡本 由起子	元東京家政学院大学	小川 恵司	凸版印刷 (株)
神立 孝一	創価大学	五島 敏芳	国文学研究資料館
阪口 哲男	筑波大学	白鳥 裕	大日本印刷 (株)
菅原 秀明	国立遺伝学研究所	太原 育夫	東京理科大学
田良島 哲	東京国立博物館	時実 象一	愛知大学
中川 優	和歌山大学	長田 孝治	(株) カテナ
長塚 隆	鶴見大学	中山 堯	神奈川大学
中山 伸一	筑波大学	西川 宜孝	みずほ情報総研 (株)
西澤 正巳	国立情報学研究所	西脇 二一	奈良大学
根岸 正光	国立情報学研究所	原 正一郎	京都大学
原田 隆史	慶應義塾大学	藤井 賢一	産業技術総合研究所
藤原 譲	筑波大学名誉教授	細野 公男	慶應義塾大学名誉教授
村川 毅彦	和歌山大学	安永 尚志	人間文化研究機構
山本 毅雄	情報学研究所名誉教授	山本 昭	愛知大学

---

---

### ■複写をされる方に

本誌に掲載された著作物を複写したい方は、(社)日本複写権センターと包括複写許諾契約を締結されている企業の従業員以外、著作権者から複写権等の行使の委託を受けている次の団体から許諾を受けて下さい。

著作物の転載、翻訳のような複写以外の許諾は、直接本会へご連絡ください。

〒107-0052 東京都港区赤坂 9-6-41 乃木坂ビル 学術著作権協会

TEL: 03-3475-5618 FAX: 03-3475-5619 E-mail: naka-atsu@muji.biglobe.ne.jp

アメリカ合衆国における複写については、次に連絡してください。

Copyright Clearance Center, Inc. 222 Rosewood Drive, Danvers, MA. 01923, USA

TEL: 978-750-8400 FAX: 978-750-4744 URL: <http://www.copyright.com/>

情報知識学会誌 Vol. 20, No. 1 2010 年 2 月 26 日発行 編集・発行情報知識学会

頒布価格 3000 円

---

---

### 情報知識学会(JSIK: Japan Society of Information and Knowledge)

会長 根岸 正光

事務局 〒110-8560 東京都台東区台東 1-5-1 凸版印刷 (株) 内

TEL: 03(3835)5692

FAX: 03(3837)0368

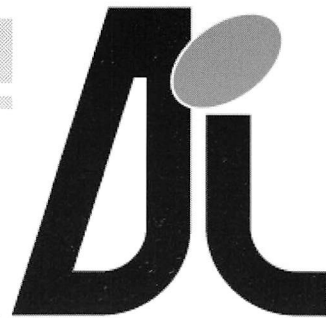
E-mail: [jsik@nifty.com](mailto:jsik@nifty.com)

URL: <http://www.jsik.jp>

---

---





ディスクロージャー・イノベーションは  
電子文書の専門家集団が運営する  
「研究開発型企业」です。

ご依頼者・ご利用者のニーズ  
WEB系アプリケーションご提供

入力・編集

蓄積

検索

表示・印刷

伝達・公証

ディスクロージャー・イノベーションが提供する  
統合的な電子文書管理操作環境

技術  
シ  
ー  
ズ

特殊ブラウザ	RDBへの文書蓄積	パトリシア方式	XML→PDF	データ配信システム
OCR技術	XMLDB	文字成分表方式	XML→HTML	電子メール自動送信
RTF経由変換	ディレクトリ構成	倒置型ファイル方式	XSL, XSLT	SOAP利用システム連携
構成管理		ベクトル検索方式		PKI技術利用
アノテーション管理		RDBとの組み合わせ		各種暗号技術
CMS				ハッシュ関数
<情報マッピング・連携・関係性定義技術> リファレンスモデル TopicMaps <基盤技術> XML関連 セマンティックWEB				

主な事業内容

情報提供サービス機構の実現

- 大学・研究機関における情報発信・研究情報提供サービス
- 研究情報基盤システムの構築ならびにドキュメント デリバリー機構の実現
- 文書類のSGML・XML・HTML化関連サービス

電子文書を利用した社会機構の調査

- 地方自治体における高度情報化計画策定調査
- Webサービス、エレクトロニックコマース、デジタルマネーの実現のための調査
- 大学研究機関におけるバーチャルシステムズの研究
- 地方自治体における高度情報化計画策定調査

電子文書関連ツール&サービス開発と販売

- (株)日本電子公証機構  
電子公証サービス dPROVE, ePROVE  
電子認証サービス iPROVE
- オントピア社 Ontopia Knowledge Suite
- イースト社他との協業機関(Synest)による、Synest Labonote (電子研究ノート)

ネットワーク関連(管理と開発)

- 研究機関におけるネットワーク構成のデザインならびに維持運用管理
- 情報提供サービスにおけるセキュリティシステムならびに課金システムの構築
- Webページの製作・構成ならびに維持
- BtoB BtoCコマース機構の開発
- コンテンツ・マネジメント構造の実現と運用

文書情報管理

- フルテキスト(全文)検索システム構築
- 検索に関わる標準プロトコルの実装機能の設計
- SGML・XML・HTML・PDF・TeX 等の文書形式相互変換システムの構築
- TopicMaps・RDF・DublinCore等のメタデータを利用した情報運用のための管理システムの構築
- 既存の紙書面による文献・資料群の電子化とデータベース化

ディスクロージャー・イノベーション株式会社  
シナジー・インキュベート事業部

171-0033 東京都豊島区高田三丁目23番10号 宝印刷本社別館(5号館クリスタルエイトビル) 5F.  
Tel : 03-5985-0920 Fax : 03-5985-0921 URL : <http://www.di-inc.co.jp/>



# *Journal of Japan Society of Information and Knowledge*

## Contents

### Research Papers

- Application of the Long Term Signature to Office Open XML files  
Kazutsuna YAMAJI, Toshiyuki KATAOKA, Naoto MIYACHI, Noboru  
SONEHARA . . . . . 1
- Automatic Term Recognition Using the Corpora of the Different Academic Areas  
Junko KUBO, Keita TSUJI, Shigeo SUGIMOTO . . . . . 15

### Investigation Report

- A Study on Recording Tools Affecting Memory and Understanding: A Comparison  
between Handwriting and Keyboard Input  
Shiro SONEHARA, Atsuko SAITO . . . . . 32

### Review

- Fair Use of Copyrighted Materials  
Soichi TOKIZANE . . . . . 38

### Research Group Report

- A Colloquium on Databases of Science and Technology: 2009 Report from the  
CODATA Activity Group  
Tetsuya BABA, Yuichiro YAMASHITA, Shuichi IWATA . . . . . 47

### News and Meetings

- 2010 Best Paper Award Ballots  
18<sup>th</sup> Annual JSIK Meeting: Call for Papers  
Changes in Membership Categories and Other Reforms  
Instructions for Authors  
News from the Secretariat

**情報知識学会誌** 第20巻1号 2010年2月26日発行

編集兼発行人 情報知識学会 〒110-8560 東京都台東区台東1-5-1 凸版印刷(株)内

TEL:03(3835)5692 FAX:03(3837)0368 E-mail:jsik@nifty.com

URL: <http://www.jsik.jp/> (振替: 00150-8-706543)

学術刊行物 ISSN0917-1436