

**知識情報処理技術に関する  
調査研究報告書**

2008年3月

社団法人 電子情報技術産業協会  
知識情報処理技術専門委員会



## 知識情報処理技術委員会 委員名簿

(敬称略・順不同)

委員長	橋田浩一	独立行政法人産業技術総合研究所
委員	井佐原均	独立行政法人情報通信研究機構
”	石川徹也	東京大学
”	辻井潤一	東京大学
”	石崎俊	慶應義塾大学
”	徳永健伸	東京工業大学
”	奥雅博	日本電信電話(株)
”	村田稔樹	沖電気工業(株)
”	福持陽士	シャープ(株)
”	木下聡	(株)東芝
”	佐藤研治	日本電気(株)
”	清野正樹	松下電器産業(株)
”	今村誠	三菱電機(株)
”	亀田雅之	(株)リコー
”	西野文人	(株)富士通研究所
”	森本康嗣	(株)日立製作所
事務局	設楽哲	(社)電子情報技術産業協会
”	北崎和枝	(社)電子情報技術産業協会

## 言語資源分科会 委員名簿

(敬称略・順不同)

委員長	井佐原 均	独立行政法人情報通信研究機構
幹事	白井 清昭	北陸先端科学技術大学院大学
委員	黒橋 禎夫	京都大学
"	井上 聡子	独立行政法人科学技術振興機構
"	斎藤 邦子	日本電信電話(株)
"	佐田 いち子	シャープ(株)
"	釜谷 聡	(株)東芝
"	中澤 聡	日本電気(株)
"	秋良 直人	(株)日立製作所
"	潮田 明	(株)富士通研究所
"	三上 崇志	三菱電機(株)
"	井田 裕子	(株)リコー
"	内元 清明	独立行政法人情報通信研究機構
"	丸山 岳彦	独立行政法人国立国語研究所
"	石川 真奈見	特定非営利活動法人言語資源協会
事務局	北崎 和枝	(社)電子情報技術産業協会

## マルチモーダルコンテンツ技術分科会 委員名簿

(敬称略・順不同)

委員長	橋田浩一	独立行政法人産業技術総合研究所
委員	石崎俊	慶應義塾大学
"	斎藤博昭	慶應義塾大学
"	加藤恒昭	東京大学
"	大森久美子	(株)NTTデータ
"	大沼宏行	沖電気工業(株)
"	野本忠司	国文学研究資料館
"	筒井秀樹	(株)東芝
"	山下信行	日本電気(株)
"	橋本三奈子	富士通(株)
"	宮原景泰	三菱電機(株)
"	伊藤一成	青山学院大学
事務局	北崎和枝	(社)電子情報技術産業協会

## 情報アクセス技術分科会 委員名簿

(敬称略・順不同)

委員長	徳永健伸	東京工業大学
幹事	藤井敦	筑波大学
委員	池野篤司	沖電気工業(株)
"	國分智晴	(株)東芝
"	神谷俊之	日本電気(株)
"	森田哲之	日本電信電話(株)
"	佐藤祐介	(株)日立製作所
"	志賀聡子	(株)富士通研究所
"	相川勇之	三菱電機(株)
"	中臣政司	(株)リコー
事務局	北崎和枝	(社)電子情報技術産業協会

# 目 次

委員名簿

目 次

エグゼクティブサマリ

1. はじめに .....	1
2. マルチモーダルコンテンツ技術 .....	3
2.1 はじめに .....	3
2.2 技術動向調査 .....	4
2.2.1 映像メタデータ活用 .....	4
2.2.2 情報家電の連携 .....	8
2.3 市場動向調査 .....	16
2.3.1 動画の構造化の種類 .....	16
2.3.2 「ニコニコ動画」 .....	16
2.3.3 「テレビブログ」 .....	17
2.4 ISO/TC37/SC4における国際標準化 .....	19
2.5 利活用技術 .....	27
2.5.1 E-learningと評価 — E-Learningの実践例の紹介 — .....	27
2.5.2 対話型学習システムを構築するツールTuTalkについて .....	31
2.6 作業研究 .....	34
2.6.1 概要 .....	34
2.6.2 データ構造化 .....	34
2.6.3 プロトタイプ .....	36
2.6.4 まとめ .....	37
3. 言語資源 .....	39
3.1 はじめに .....	39
3.2 活動報告 .....	40
3.2.1 言語情報処理ポータル .....	40

3.2.2	言語資源カタログの整備	42
3.2.3	ヒアリング調査報告I：KOTONOHA計画における著作権処理	49
3.2.4	ヒアリング調査報告II：日外アソシエーツにおける言語資源配布の実態	54
3.3	今後の展望	58
4.	情報アクセス技術	59
4.1	はじめに	59
4.2	行動履歴情報の利用に関する技術動向	61
4.2.1	行動ターゲティング広告	61
4.2.2	推薦システム	68
4.2.3	情動分析	70
4.2.4	センサネットワーク	73
4.2.5	ホームネットワーク	75
4.3	履歴情報を活用したサービス事例	88
4.3.1	調査の概要	88
4.3.2	調査結果	90
4.4	推薦とパーソナライゼーションの研究動向	99
4.4.1	Recommender Systems in E-Commerce	99
4.4.2	Analysis of Recommendation Algorithms for E-Commerce	101
4.4.3	Enabling Scalable Online Personalization on the Web	103
4.4.4	Red Opal:Product-Feature Scoring from Reviews	107
4.4.5	Optimal Pricing with Recommender Systems	110
4.4.6	Recommender Systems and their Impact on Sales Diversity	111
4.4.7	Recommending or Persuading? The Impact of Shopping Agent's Algorithm on Use Behavior	114
4.5	SIGIR2007参加報告	118
4.5.1	はじめに	118
4.5.2	チュートリアル	118
4.5.3	本会議	119
4.5.4	Karen Sparc Jones氏のビデオ講演	120
4.5.5	おわりに	121

## エグゼクティブサマリ

知識管理、テキストマイニング、情報検索など、組織や個人の情報管理・発信や知識創造の効率化に資する情報技術は、急速に進展する社会の情報化・知識化において不可欠な基盤となっている。高度な情報通信ネットワーク社会の出現は、人間の生活世界を構造化し、時空を越えた情報や知識の取得と共有と循環を可能にすることで、個人の生活や組織の業務を急速に変えつつある。このような「知識に基づく社会」において個人と組織が有用な知識を容易に取得・創造・発信・共有し拡大再生産するための知識情報処理技術が、本調査研究の対象である。

知識情報処理技術専門委員会においては、自然言語処理や情報アクセス技術等の知識情報処理技術およびそれに関連するデジタルコンテンツやビジネスの動向に関して調査研究を進めている。平成19年度は、マルチモーダルコンテンツ技術分科会、情報アクセス技術分科会、言語資源技術分科会という3つの分科会の活動を進めるとともに、知識情報処理技術シンポジウム「知識に基づくマーケティング～量から質へ～」を開催し、また知識情報処理技術に関する市場調査を行なった。

マルチモーダルコンテンツ技術分科会では、映像コンテンツ等の産業において要請が高まりつつある新たなビジネスモデルの基盤となる技術の探究を促進すべく、意味的な構造を明示することによって映像等のコンテンツの作成を効率化する技術、そのような仕方で作られたマルチモーダルコンテンツをさまざまな形でインタラクティブに高度利用する方法、そうした利用法に基づくビジネスモデル等について検討している。これまでに2種類の対話課題を収録したマルチモーダル対話コーパスを製作し、それに音声韻律タグ、構文・意味情報を明示するGDAタグ、発話の役割を示す談話タグなどさまざまなアノテーションを付与して、ビデオデータとともに「JEITAマルチモーダル対話コーパス」として言語資源協会（GSK）から配布されている。

平成19年度は以上のような作業や調査から得た知見をもとに、これからますます多く流通するであろうマルチメディアコンテンツを高度利用するための技術の現状を調査し可能性を探った。まず、映像メタデータの活用と情報家電との連携等の視点から、技術動向および市場動向の調査を行った。また、言語コンテンツへのさまざまなアノテーションの方式に関するISO/TC37/SC4の活動に参画して国際標準化の動向を調査するとともに、談話関係（文間の関係）のデータカテゴリの策定等に貢献した。eラーニングに関しては、本分科会ではこれまでマルチメディアコンテンツの一例としてeラーニング教材に関する技術を取り上げてきたが、平成19年度はeラーニングにおける最近の技術動向やeラーニングと自然言語処理技術の関わりに関する調査を進めた。一方、数年来開発を進めているマルチモーダルコンテンツの構造化オーサリングツールについては、ビデオコンテンツをオントロジーに基

づく木構造によって意味的に構造化する機能を加え、それを要約して表示する機能を開発した。

**情報アクセス技術分科会**では、システムやサービスなどの履歴情報の収集とその利用技術に関する動向調査を行った。すべてのユーザに同一のインタフェースやサービスを提供するのではなく、各ユーザの性質に合わせたきめ細かいサービスを提供する技術に関心が集まっている。このためにはユーザの利用履歴などからユーザ・モデルを構築する技術が重要となる。また、セキュリティの観点から各ユーザのシステムの利用履歴を保存し、それを監視することにより警告を発したり、問題が発生したときには分析して対応策を講じたりするなどの利用法が考えられる。

履歴情報を利用するアプリケーションと履歴情報を収集するためのセンシング技術という2つの観点からヒアリングとショールームの見学による7件の調査をおこなった。これらは、履歴情報の利用に関する事例、履歴情報を収集する技術、ホームネットワークの領域における実装技術・標準化技術とそれを実装して利用した具体例に及ぶ。2007年度の国内広告費では、ついにインターネット広告が雑誌広告を上回ったとされており、中でもユーザの行動履歴からユーザの興味を推定し、その興味に合致した広告をそのユーザに対して集中的にうつという「行動ターゲティング広告」も注目を集めている。今後、さらに新しいタイプのインターネット広告が現れる可能性もある。

インターネットを検索することにより、履歴情報を利用したサービスを網羅性に収集し、それをいくつかの観点から分類した。収集した事例を分析し、分類の軸として、Web上の履歴情報か実世界の履歴情報という対立軸、履歴情報を分析した2次利用かそのままの利用かという対立軸、何に対する行動の履歴であるかという対象の軸などを利用した。

また、この分野の研究動向を調査するために、ACM Conference on Electronic Commerceで発表された論文から、推薦やパーソナライズに関連する7件、および情報検索関連の著名な国際会議SIGIR2007の概要と今年度の調査項目に関連の深いいくつかの発表の概要を報告する。

**言語資源技術分科会**では、言語資源の研究開発利用に関わる問題点を調査することを目的としている。著作権に関する調査、言語資源カタログの整備、また言語資源に限らず広く言語処理に関する情報提供の場としての言語情報処理ポータルサイトの運営を、言語資源協会（GSK）と協調しながら行っている。

平成19年度はこれまでの成果に基づいて言語情報処理ポータルサイトの運営をさらに進めた。会議案内、製品ニュース、コラムの定期的な更新と共に、その他のコンテンツについても迅速な更新を行っている。当サイトは言語情報処理に関連する情報を網羅的・総合的に提供しており、当該分野および関連分野の研究者、開発者、学生などから好評を得ている。新しい試みとして、会議案内の欄で、メジャーなカンファレンスに対してのWikiページを作成し、皆で情報の更新や共有が可能な場を提供する試みを実施した。

また、言語資源カタログの整備を行った。この作業は情報通信研究機構や言語資源協会と協力しながら行った。情報通信研究機構では、名古屋大学と共同で、2007年から大規模言語資源メタデータベースSHACHIの構築を開始している。SHACHIは、日本・アジア諸国の言語資源をはじめ、世界中の言語資源の詳細なメタデータを収集し、得られた知見を基に体系的な蓄積を試みている。すでに2000件を超える言語資源が収録され、検索機能も装備している。SHACHIはすでに、<http://www.shachi.org> で公開しており、言語処理ポータルからもリンクが張られている。

言語資源についての幅広い知見を収集するため、有識者のヒアリングを行った。国立国語研究所の前川喜久雄氏にKOTONOHA計画における著作権処理について講演していただいた。日外アソシエーツの星俊雄氏に言語資源配布の実態について講演していただいた。このヒアリングをきっかけに、スポーツ報知新聞社のデータが研究利用可能となった。

**知的情報技術に関する市場動向調査**として、平成19年度には、Webサービスの利用状況に関するアンケート調査を行った。検索、旅行、地図、経路情報サービス、グルメガイド、映像サービス、翻訳、辞書・事典等のサービスに関してそれぞれいくつかの代表的なサイトを取り上げ、性別と年齢（15～19歳、20代、…、60代）の組み合わせごとに50～100人程度の回答者（全部で約1700人）から情報を収集し、これを分析した。



## 1. はじめに

知識情報処理技術専門委員会においては、自然言語処理や情報アクセス技術等の知識情報処理技術およびそれに関連するデジタルコンテンツやビジネスの動向に関して調査研究を進めている。平成19年度は、マルチモーダルコンテンツ技術分科会、情報アクセス技術分科会、言語資源技術分科会という3つの分科会の活動を進めるとともに、知識情報処理技術シンポジウム「知識に基づくマーケティング～量から質へ～」を開催し、また知識情報処理技術に関する市場調査を行なった。

マルチモーダルコンテンツ技術分科会では、映像コンテンツ等の産業において要請が高まりつつある新たなビジネスモデルの基盤となる技術の探究を促進すべく、意味的な構造を明示することによって映像等のコンテンツの作成を効率化する技術、そのような仕方で作られたマルチモーダルコンテンツをさまざまな形でインタラクティブに高度利用する方法、そうした利用法に基づくビジネスモデル等について検討している。

平成19年度は以上のような作業や調査から得た知見をもとに、これからますます多く流通するであろうマルチメディアコンテンツを高度利用するための技術の現状を調査し可能性を探った。まず、映像メタデータの活用と情報家電との連携等に関して技術動向および市場動向の調査を行った。また、言語コンテンツへのさまざまなアノテーションの方式に関するISO/TC37/SC4の活動に参画した。eラーニングに関しては、最近の技術動向やeラーニングと自然言語処理技術の関わりに関する調査を進めた。マルチモーダルコンテンツの構造化オーサリングツールについては、ビデオコンテンツを木構造によって構造化し要約する機能を開発した。

情報アクセス技術分科会では、システムやサービスなどの履歴情報の収集とその利用技術に関する動向調査を行った。

履歴情報を利用するアプリケーションと履歴情報を収集するためのセンシング技術という2つの観点からヒアリングとショールームの見学による調査をおこなった。これらは、履歴情報の利用に関する事例、履歴情報を収集する技術、ホームネットワークの領域における実装技術・標準化技術とそれを実装して利用した具体例に及ぶ。

インターネットを検索することにより、履歴情報を利用したサービスを網羅的に収集し、それをいくつかの観点から分類した。収集した事例を分析し、分類の軸として、Web上の履歴情報か実世界の履歴情報という対立軸、履歴情報を分析した2次利用かそのままの利用かという対立軸、何に対する行動の履歴であるかという対象の軸などを利用した。

また、ACM Conference on Electronic Commerceで発表された推薦やパーソナライズに関連する論文、および情報検索の国際会議SIGIR2007の概要と今年度の調査項目に関連の深いいくつかの発表の概要をまとめた。

言語資源技術分科会では、言語資源の研究開発利用に関わる問題点を調査することを目的としている。著作権に関する調査、言語資源カタログの整備、また言語資源に限らず広く言語処理に関する情報提供の場としての言語情報処理ポータルを、言語資源協会（GSK）と協調しながら行っている。

平成19年度はこれまでの成果に基づいて言語情報処理ポータルをさらに進めた。会議案内、製品ニュース、コラムの定期的な更新と共に、その他のコンテンツについても迅速な更新を行っている。会議案内の欄で、メジャーなカンファレンスに対してのWikiページを作成し、皆で情報の更新や共有が可能な場を提供する試みを実施した。また、情報通信研究機構や言語資源協会と協力しながら言語資源カタログの整備を行った。さらに、言語資源についての幅広い知見を収集するため、有識者のヒアリングを行った。国立国語研究所の前川喜久雄氏にKOTONOHA計画における著作権処理についてご講演いただいた。日外アソシエーツの星俊雄氏に言語資源配布の実態について講演していただいた。このヒアリングをきっかけに、スポーツ報知新聞社のデータが研究利用可能となった。

知的情報技術に関する市場動向調査として、平成19年度には、Webサービスの利用状況に関するアンケート調査を行った。検索、旅行、地図、経路情報サービス、グルメガイド、映像サービス、翻訳、辞書・事典等のサービスに関してそれぞれいくつかの代表的なサイトを取り上げ、性別と年齢（15～19歳、20代、…、60代）の組み合わせごとに50～100人程度の回答者（全部で約1700人）から情報を収集し、これを分析した。

## 2. マルチモーダルコンテンツ技術

### 2.1 はじめに

マルチモーダルコンテンツ技術分科会は、対話コンテンツをはじめとする様々なコンテンツの処理・利用技術の調査・発展に資することを目的として活動している。コンテンツとしては、言語情報が主に入っている映像データを取り上げ、自然言語処理を用いた言語解析はもちろんのこと、韻律、動作、表情といったモダリティについても分析を行なっている。このようなさまざまな解析結果をアノテーションとして付与することで、映像データが構造化され、その構造を利用して検索や要約といったより知的で包括的なコンテンツの利用が可能になると考える。

本委員会では、これまでに2種類の対話課題を収録したマルチモーダル対話コーパスを製作し、それに音声韻律タグ、構文・意味情報を明示するGDAタグ、発話の役割を示す談話タグなどさまざまなアノテーションを付与してきた。製作したアノテーションデータはビデオデータとともにJEITAマルチモーダル対話コーパスとして言語資源協会から有償配布されている。入手方法等については言語資源協会のホームページ<http://www.gsk.or.jp/> をごらんいただきたい。

今年度はそのような具体的な作業から得た知見をもとに、これからますます多く流通するであろうマルチメディアコンテンツを高度利用するための処理技術の現状と可能性を探った。まず、2.2節で映像メタデータの活用と情報家電との連携といった視点から技術動向調査の結果を述べる。2.3節では、そのような技術を受け入れる市場動向について述べる。

コンテンツへのアノテーション技術の標準化動向として、ISOにおける談話関係（修辞構造）アノテーションを取り上げ、2.4節で詳述する。

本委員会ではこれまでマルチメディアコンテンツの一例としてeラーニング教材を取り上げてきたが、eラーニングにおける最近の技術動向やeラーニングと自然言語処理技術の関わりについて2.5節で取り上げる。

数年来開発を進めているコンテンツの構造化オーサリングツールについては、2.6節で今年度の進捗状況を述べる。

蓄積されるコンテンツが質を変えながらその量を加速度的に増やしている現在、それをいかに利用するかがますます重要になっている。当委員会への要望、意見などお寄せいただければ幸いである。

(斎藤博昭)

## 2.2 技術動向調査

### 2.2.1 映像メタデータ活用

本項では、映像を対象としたメタデータ活用の技術動向を報告する。映像のメタデータということでは、YouTube<sup>1)</sup>に代表される動画共有サービスの成長が著しいが、ここでは、統制のとれた（所定ルールに従った）メタデータを対象とし、映像編集と番組映像視聴におけるメタデータ活用例を示す。

#### 2.2.1.1 映像編集におけるメタデータ活用例

映像編集における活用例として、アップル社の映像編集ソフトFinal Cut Pro<sup>2)</sup>で用いられているFinal Cut Pro XML Interchange Format<sup>3)</sup>について説明する。同ソフトでは、素材データや編集データなどの様々な情報をXML出力することで、プロジェクト（素材や編集、設定の管理情報）の詳細を他のアプリケーションに転送可能であり、また、テキスト編集アプリケーションなどで編集されたXMLファイルを読み込んで、プロジェクトを更新することも可能となっている。このフォーマットは、2003年に策定された後、2006年にバージョン3、2007年にFinal Cut Pro 6.0対応でバージョン4がリリースされ、アップデートが続けられている。

Final Cut Proにおいて、最も基本的なオブジェクトはクリップ（メディアを仮想的に表現するもの）であり、このクリップのコード化例を図2.2.1-1に示す。

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE xmeml>
<xmeml version="4">
  <clip>
    <name>Traffic</name>
    <duration>264</duration>
    <rate>
      <ntsc>TRUE</ntsc>
      <timebase>30</timebase>
    </rate>
  </clip>
</xmeml>
```

図2.2.1-1 クリップのコード化例

最初の3行は本フォーマット共通であり、1行目はXML宣言、2行目はDOCTYPE宣言で、3行目のxmeml要素は本フォーマットのルート要素である。クリップをコード化するclip要素には、以下の三つの要素が含まれている。

- Name 要素 . . . クリップの名前を表す
- duration 要素 . . . クリップ内の総フレーム数を表す
- rate 要素 . . . クリップのフレームレート（再生速度）を表す

rate 要素には、timebase要素とntsc要素の二つの部分要素がある。Timebase要素はフレームレートの基準値を表すものである。Ntsc要素は、timebase要素の値をそのまま用いるか、あるいはNTSC方式と同様に0.1%低下させて用いるかを決定する。図2.2.1-1の例では、フレームレートはNTSC方式と同じ29.97フレーム／秒となる。

さらに本フォーマットでは、複数クリップを含むシーケンスを表したり、トランジション（カット遷移時の特殊効果）やフィルタ処理などのエフェクト（特殊効果）を表すこともできる。表2.2.1-1～表2.2.1-2に本フォーマットの要素一覧を示す。バージョン3からは、metadata要素とその部分要素（表2.2.1-2の分類10）を導入し、QuickTimeのメタデータにも対応している。

表2.2.1-1 要素一覧1

分類	要素	機能
1) 主要要素	xmeme project bin children sequence media track locked outputchannelindex link linkclipref clipindex groupindex uuid updatebehavior itemhistory	ルート要素やプロジェクトなどの管理情報、クリップ間のリンク情報など
2) クリップ	clip clipitem file pathurl mediaSource anamorphic alphatype alphareverse compositemode masterclipid ismasterclip logginginfo description scene shottake lognote good scenenote shotnote takenote labels label label2 comments mastercomment[n] clipcommenta clipcommentb sourcetrack start end subclipinfo startoffset endoffset stillframe stillframeoffset syncoffset multiclip angle activevideoangle activeaudioangle collapsed synctype mediadelay defaultangle subframeoffset mixedratesoffset gamma	クリップで用いるメディアファイルの情報や、ログ記録、マルチクリップ情報など、クリップに関わる情報を表す
3) ビデオと オーディオ	video layerindex audio channelcount trackcount channeldescription layout audiochannel channellabel sourcechannel outputs group index numchannels downmix channel format samplecharacteristics width height pixelaspectratio fielddominance colordepth codec depth samplerate	ビデオメディアやオーディオメディア特有のデータを表す

表2.2.1-1 要素一覧2

分類	要素	機能
4) 共通要素	name duration enabled marker comment in out mediatype trackindex	名前やフレーム数など、共通的な情報を表す
5) 速度とタイムコード	rate timebase ntsc timecode string frame displayformat field reel source	再生速度やタイムコードなどを表す
6) エフェクト	generatoritem primarytimecode transitionitem alignment filter effect effecteffect effectcategory effecttype effectclass wipecode wipeaccuracy startratio endratio reverse privatestate parameter parameterid valuemin valuemax valuelist valueentry value keyframe when inscale inbez outscale outbez red green blue alpha horiz vert interpolation	トランジションやフィルタエフェクトなど、エフェクトに関連する情報を表す
7) タイムリマップエフェクト	speedvirtualkf speedkfin speedkfout speedkfstart speedkfbend origvalue hadbezierin hadbezierout	タイムリマップエフェクト(再生速度・方向を制御することによるエフェクト)のためのキーフレーム情報をコード化する
8) アプリケーション特有データ	appspecificdata appname appmanufacturer appversion data fcpimageprocessing useyuv usesuperwhite rendermode qtcodec codecname codectypename codectypecode codecvendorcode spatialquality temporalquality keyframerate datarate qteffectid UUID	アプリケーション名やバージョン番号、コーデックや画像処理に関する情報など、アプリケーション特有のデータを表す
9) フィルムデータ	filmdata filmslate scene take slate cameraroll labroll dailyroll keycode prefix feetframes perffoffset reversed inknumber filmstandard telecinespeed	シーンやテイクの情報など、Cinema Tools(フィルム編集用ソフト)から取り込まれたメタデータを表す
10) QuickTimeメタデータ	metadata storage key size type value	QuickTimeファイル中のメタデータを表す
11) インポートオプション	importoptions createnewproject targetprojectname defsequencepresetname filterreconnectmediafiles filterincludemarkers filterincludeeffects filterincludesequencesettings displaynonfatalerrors createfcpprojectatxmlfilepath deletethisxmlfileatimport	Final Cut Proに特有のインポートオプションを表す

### 2.2.1.2 番組映像視聴におけるメタデータ活用例

番組映像を視聴する際に、メタデータを使って手軽に検索・再生できる事例として、携帯電話向けアプリケーション「Mojie」と、動画管理・再生ソフト「Tagiri」を紹介する。

### 2.2.1.3 Mojie

2007年10月11日の日本放送協会（以下、「NHK」）のプレスリリース<sup>4)</sup>によると、NHKと株式会社インデックス（以下、「インデックス」）は携帯電話向けのビデオオンデマンドアプリケーション「Mojie」を共同開発し、これが同月フランスで開催中の映像コンテンツ見本市MIPCOMにおいて“Mobile & Internet TV Awards”審査員奨励賞を受賞した、とのことである。本アプリケーションは、テレビ番組の字幕放送用のテキストデータを活用したもので、以下の特徴を持つ。

### 2.2.1.4 検索機能

テレビ番組の字幕テキストをメタデータとして活用することで、いろいろなキーワードでシーン単位の検索を行なうことができる。

#### 1) 「読むモード」機能

検索機能で選択した番組を、字幕テキストと静止画像（字幕に対応するサムネイル画像）により、電子ブックのように“読む”ことができる。

#### 2) 「観るモード」機能

「読むモード」で読んでいる時に字幕テキストをクリックすれば、対応する動画を視聴できる。さらに、特定シーンを友人に勧めたり、アクセスの多いシーンをダイジェストとして視聴することも可能である。

なお、インデックスの発表<sup>5)</sup>によれば、2007年12月10日から、インデックスの運営するモバイルサイト「速報Musicサーチ」の「速M.tv（ソクエムドットティービー）」コーナーにて、「Mojie」を使った動画コンテンツ配信サービスが開始され、NHKのコンテンツ「みんなのうた」などが提供されている。

### 2.2.1.5 Tagiri

Tagiri<sup>6)</sup>とは、「テレビブログ」<sup>7)</sup>（テレビ番組をテーマにしたブログサービス）を運営している株式会社メタキャストが無料配布している動画管理・再生ソフトである。「テレビブログ」のサイトからテ

レビ番組のチャプター情報（番組内の詳細な放送内容と対応する時間を示すデータ）を自動・手動でダウンロードし、PC内にある録画映像に付与することができる。自動ダウンロードでは、所定のフォルダにある映像に対し、ファイル名や更新日時などから、どの番組を録画したものか自動で判断して、該当する番組チャプターを取得する。任意のシーンをチャプター情報として設定することも可能であり、映像の詳しい内容をキーワード検索できるので、探しているシーンに素早くアクセスできる。なお、2008年2月時点で、番組チャプターを取得できるのは、関東エリアの以下の放送局の番組となっている。

NHK総合・NHK教育・日本テレビ・TBSテレビ・フジテレビ・テレビ朝日・テレビ東京

## 参考文献

- 1) <http://www.youtube.com/>
- 2) Apple Inc. : Final Cut Pro 6 ユーザーズマニュアル
- 3) Apple Inc. : Final Cut Pro XML Interchange Format
- 4) <http://www3.nhk.or.jp/pr/marukaji/m-giju188.html>
- 5) [http://www.indexweb.co.jp/2007/12/20071210\\_1400.html](http://www.indexweb.co.jp/2007/12/20071210_1400.html)
- 6) <http://tagiri.jp/>
- 7) <http://www.tvblog.jp/>

(宮原景泰)

### 2.2.2 情報家電の連携

近年、情報家電は、様々な機器がつながるようになり、複雑化・高度化が進み、使い方が難しくなってきた。さらに、ユーザの求める機能を複数の機器が連携して実現することも可能になってきており、また、同じ機能を複数の方法で実現することも可能になっている。

例えば、テレビ番組をハードディスクレコーダに保存するとしても、圧縮のレベルを変更できたり、2番組を同時に録画できる。また、ハードディスクレコーダの録画予約状況やディスクの空き容量など、その時々状況に応じて、ユーザが利用したい機能も様々である。

しかしながら、これらの機能をユーザが使いこなし、適切に利用できるとは限らない。そこで、情報家電などの機器が、ユーザの状況に応じて、適切に制御しあったり連動して動作してユーザをナビゲートするなど、ユーザを支援できると便利である。

本節では、情報家電のサービス連携を行うための技術動向について述べる。最初の2つの論文は、情報家電の連携に必要なと思われるエージェント技術や知識表現に関する取組みである。次の論文は、ユーザ嗜好を考慮した情報家電の連携に関するものである。最後の論文は、情報家電に対する

操作と状態変化を考慮した因果関係知識を利用した研究である。これらの実用化はしばらく先と考えられるが、今後のホームネットワーク環境で有効な技術であると考えられる。

### 情報家電エージェント協調のためのオントロジーサービスの実現 [1]

この論文では、情報家電の相互協調動作を実現するためにエージェントを使用し、各情報家電にエージェントを配置して協調動作を行わせることを提案している。この枠組みでは、機器間の協調動作のためのやりとりをいくつかの典型的なパターンに分類し規定したエージェントの相互作用プロトコルを用いている。

また、機器オントロジー・機能オントロジーの2つのオントロジーを用いることにより、新たな機器の追加を少ない手間です適切に行うことを可能にし、協調動作時にはオントロジーを利用することで機器の柔軟な連携を可能にしている。

機器オントロジーは、図2.2.2-1のように、機器の種類を体系化したものであり、最下位のノードは、機器個別のクラスを記述している。一方、機能オントロジー[2]は、図2.2.2-2のように、機器のもつ各種機能を体系化したものであり、環境への作用、利用者への作用、対象物への作用などの抽象化がなされている。最下位のノードは、各種機能が持つ機能や操作を表している。

また、家オントロジーを規定し、機器が存在する部屋、人がいる部屋、部屋同士の位置関係なども表現している。これらのオントロジーは、ウェブ・オントロジー言語OWL[3]を用いて記述されている。

協調動作の利用例として、次のものが挙げられている。

- 1) 複数の機器が、機器間での競合を解消するために、機器間のリソースを分け合う動作
  - ・ 電話機の発信・着信時、テレビやステレオ機器等が自身の音量を下げる。
  - ・ 書斎で人が読書しているとき、書斎に隣接する部屋のAV機器の音量を下げる。
  - ・ 電力が足りない場合、足りない電力を他の機器が分け合う。
  - ・ 人がいなくなったとき、機器の使用電力を減らす。
- 2) ある機器が自身の目的を達成するために他機器の協力を要請して行う動作
  - ・ ビープ音などでしかユーザに作業完了等を通知できない機器が、テレビなどの画面表示を用いてメッセージを通知する。

機能オントロジーでは、環境への作用、利用者への作用、対象物への作用の抽象化がなされている。これによって、複数の機器が同じリソース（部屋などの環境、利用者や対象物）へ作用するものならば、お互い協調しあう必要があるということが推論できると思われる。その後、協調しあうべき機器間で、エージェント機能を用いて協調動作を実現できると推測される。

また、挙げられた動作例では、実世界の人の状況を収集するための課題もあるが、音量や電力量な

どが制御対象に挙げられており比較的シンプルなりソースを分け合っている。一方、多機能なハードディスクレコーダを制御したりするためには、さらに複雑な機能オントロジーを用意する必要があると考えられる。番組予約状況やディスクの空き容量などのリソースの詳細な表現も必要になってくる。

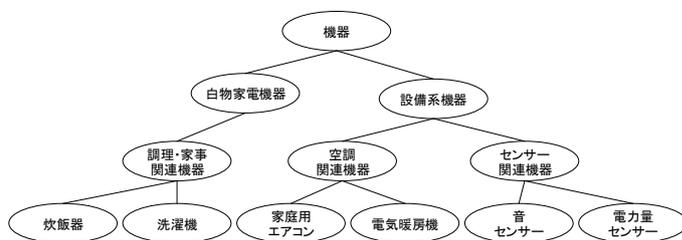


図2.2.2-1 機器オントロジーの一部

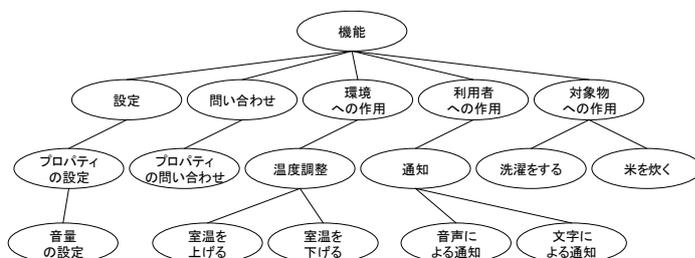


図2.2.2-2 機能オントロジーの一部

#### 情報家電オントロジー記述ガイドライン [4]

情報家電オントロジー記述ガイドラインは、情報家電サービス基盤フォーラム[5]で策定している、情報家電オントロジーを追加・記述するためのガイドラインを示したものである。

情報家電オントロジーとは、情報家電を活用・利用するにあたり必要となる情報を容易に検索できるよう、ウェブページに与えるメタデータを記述するためのオントロジーであり、ウェブ・オントロジー言語OWL[3]を用いて記述している。

情報家電オントロジー記述ガイドラインでは、メーカーや機種に依存しない「情報家電共通オントロジー」のうち、基本的な役割を果たす語彙（コア語彙）を定義している。

情報家電オントロジー記述ガイドラインで規定している主なコア語彙を、表2.2.2-1、表2.2.2-2に示す。また、コア語彙間の関係を図2.2.2-3に示す。タスクやユーザによる操作、機器の動作などを表すクラスが用意されている。DeviceクラスやFunctionクラスは、文献[1][2]の機器オントロジーや機能オントロジーと対応づけ可能だと推測できる。

また、theme、instruction、destinationなどの関係記述子が用意され、リソースに対する影響関係

なども記述できる。

コア語彙の下位概念として、様々な一般語彙を作成することで、スペック情報や家電間の接続可能性情報、操作手順等が記述可能になる。コア語彙の利用例として、「DVD録画機能は、DVDレコーダにより実行される録画機能の一種であり、映像コンテンツをDVDに保存することを特徴とする。」という記述は、図2.2.2-4のように記述可能であることが示されている。

コア語彙から派生した一般語彙を利用することで、個々の機能が影響を与えるリソースが何であるか、など様々な関係が記述可能になると推測される。

表2. 2. 2-1 情報家電オントロジーコア語彙の主なクラス

クラス名	意味
Device	機器
Function	機器の機能
State	状態
StateVariable	状態変数
Transition	状態の変化
UserOperation	ユーザによる操作
MachineAction	機器の動作
Task	タスク (=ユーザが達成しようとする目的)

表2. 2. 2-2 情報家電オントロジーコア語彙の主なプロパティ

プロパティ	意味
hasFunction	主語のリソースは目的語である機能を機能として持つ。
Causes	主語の機能の実行により目的語の事象が起きる。
hasStateVariable	主語のリソースは目的語である状態変数を状態変数として持つ。
hasPossibleStateValue	目的語である状態は、主語の状態変数の値として可能な状態である。
hasStateValue	目的語である状態は主語の状態変数の値である。
changesFrom	主語は目的語である状態からの状態変化である。
changesTo	主語は目的語である状態への状態変化である。
Affects	主語の操作は、目的語である機能の実行に影響を与える。
isRealisedBy	主語のタスクは、目的語である機能により実現される。
Theme	主語が機能、操作、動作、タスクの場合、目的語は、それらにより変化を受けるものである。 主語が状態の場合、変化を受ける状態にあるものがある場合は、目的語は、そのものである。
Instruction	主語が機能、動作、タスクの場合、目的語は、主語の実行に使われる機器である。 主語が状態の場合、目的語は、主語の実行に使われている機器である。

Destination	主語は、移動を表す機能、状態、操作、動作、タスク。目的語は、移動されるものの移動先。移動されるものはtheme、移動元はsource、移動経路はviaのそれぞれを目的語として表す。
-------------	--

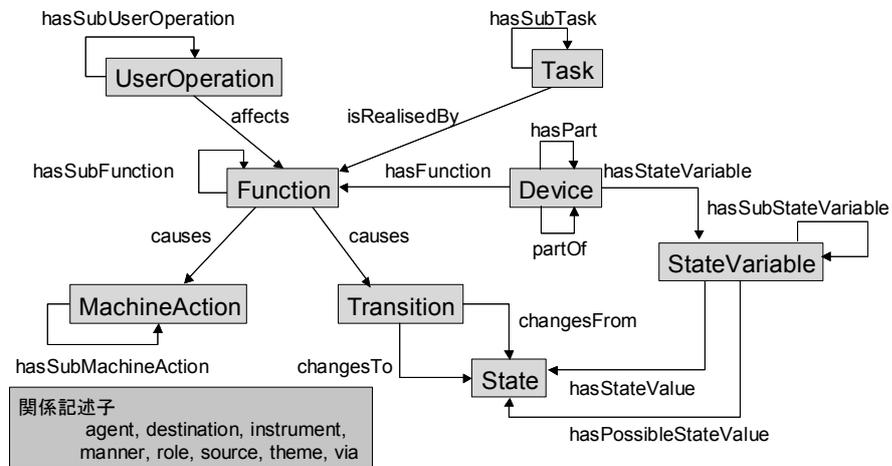


図2. 2. 2-3 情報家電オントロジーのコア語彙間の関係

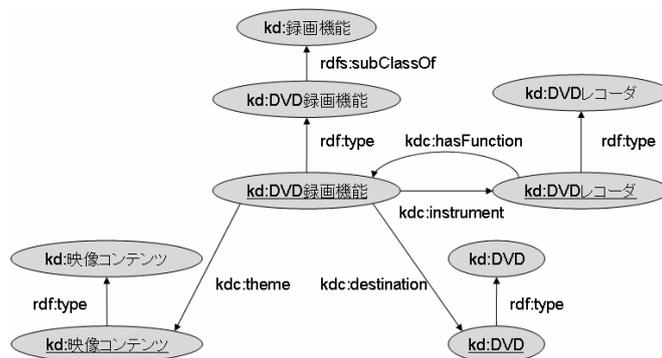


図2. 2. 2-4 DVD録画機能に関するインスタンスレベルのグラフ

### ユーザの嗜好を利用した機器自動制御支援機構 [6]

この論文では、ユーザの嗜好を反映したアクチュエータネットワークの利用支援を行う User Preference supported Architecture for device automation (UPA) を提案している。

アクチュエータとは、仮想的な環境情報を物理的な環境に出力する機器と定義している。また、センサは、物理的な環境の情報を仮想的な環境に入力する機器と定義している。例えば、エアコンは、温度を取得するセンサと温度を調節するアクチュエータの2つの機器を組み合わせたものとする。

アクチュエータの影響範囲は、閉じた空間内に制限され、アクチュエータには、機能の競合が起こ

る。例えば複数のアプリケーションが同時に音声出力を行うアクチュエータを利用した場合、音による競合が発生する。そのため、競合する機能を分類し、類似機能を提供するアクチュエータをグループ化し統合的に管理している。

アクチュエータの最も基本的なグループは、「音の出力」「映像の出力」「空間内の温度の調節」「空間内の明るさの調節」の4つのグループに分類している。

また、ユーザの好みとは、特定の事項に複数の選択肢が存在する際に、ユーザが行う選択傾向であり、状況ごとに変化するものでなく、傾向であるという点で静的な状態であるとしている。

ユーザの好みは、その対象を考慮して2つに分類している。

- コンテンツに対する好み

ここでいうコンテンツとは、映像、音楽、文書など価値が付加されたデータである。コンテンツに対する好みは一つの要素に基づく選択傾向ではなく、複数の要素に基づく選択傾向であり、定義するのが困難と予想している。

- サービスが適用する機能に対する好み

ソフトウェアが提供するものとアクチュエータが提供するものがある。サービスの機能に対する選択傾向はコンテンツに対する選択傾向に比べて少なく、具体的な要素に基づく傾向があるとしている。

サービスが適用する機能に対する好みは、例えば、テレビ番組を、ハードディスクレコーダに保存する際の圧縮のレベルの傾向や、2番組を同時に録画したい場合の方法など、複数の選択肢がある場合に、有効な考え方であると思われる。

## 情報家電の操作のための対話インタフェースの開発 [7]

この論文は、情報家電の操作のための日本語対話インタフェースの開発に関するものであり、対話処理部として、図2.2.2-5に示す構成をとっている。家電間の連携を扱ったものではないが、情報家電の状態表現バッファに情報家電の現在の状態を格納し、因果関係知識バッファに情報家電の操作に関する因果関係知識群を格納している。因果関係知識は、

- 機器に対する各操作とそれによって引き起こされる状態変化
- 状態変化が引き起こす次の状態変化
- 周辺の状態とユーザの感覚的状态を結ぶ因果関係

を記述している。

図2.2.2-6にテレビの因果関係知識の例を示す。図2.2.2-6では、テレビの電源が入っていない状態で「8チャンネルに設定して」という入力があった場合、「テレビの電源をONにする」「チャンネルを8に設定する」という2つの操作が必要になることを示している。

この例では、単一の家電機器で因果関係知識を記述しているが、複数の機器が協調する場合にも応用可能であり、これらの因果関係知識がオントロジーなどのメーカ共通の語彙で記述できると便利だと思われる。

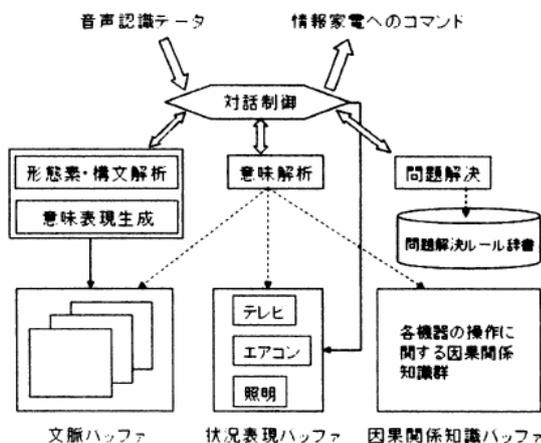


図2. 2. 2-5 対話処理部の構成図

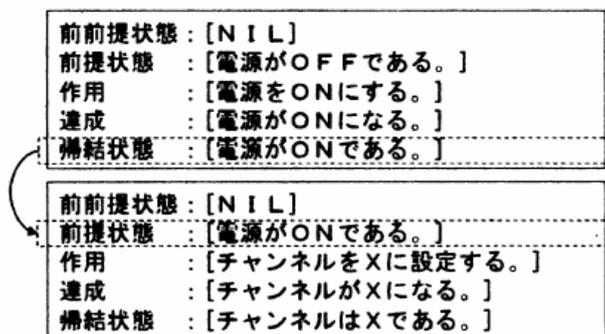


図2. 2. 2-6 テレビの因果関係知識

[参考文献]

- [1] 平尾郁実, 飯島正, 山口高平:「情報家電エージェント協調のためのオントロジーサービスの実現」, 第21回人工知能学会全国大会 (2007) .
- [2] 山田知秀, 飯島正, 山口高平:「オントロジーを利用した情報家電エージェント協調アーキテクチャ」, 第19回人工知能学会全国大会 (2005) .
- [3] OWL Web Ontology Language Guide, W3C Recommendation 10 Feb 2004. Smith, Welty, McGuinness, eds. <http://www.w3.org/TR/owl-guide/>
- [4] 情報家電オントロジー記述ガイドライン

[http://purl.oclc.org/NET/spia\\_ontology/writing/1.0](http://purl.oclc.org/NET/spia_ontology/writing/1.0)

- [5] 情報家電サービス基盤フォーラム <http://net2.intap.or.jp/SPIA/>
- [6] 松宮健太, 岩井将行, 中澤仁, 徳田英幸:「ユーザの嗜好を利用した機器自動制御支援機構」, ユーザの嗜好を利用した機器自動制御支援機構, OS, Vol.2001, No. 65, pp.89-96 (2001)
- [7] 内田尚和, 常盤大樹, 西末衣, 高木朗, 麻生英樹, 橋本政朋, 森彰, 中島秀之, 伊東幸宏, 小林一郎, 八名和夫:「情報家電の操作のための対話インタフェースの開発」, 情報処理学会研究報告, SLP, Vol.2004, No.131, pp.265-270 (2004).

(大沼宏行)

## 2.3 市場動向調査

本節では、動画の構造化に関する市場動向の調査結果を報告する。

### 2.3.1 動画の構造化の種類

動画の構造化とは、映像コンテンツにメタデータあるいはアノテーション(注釈)を付与することである。映像コンテンツにメタデータを付与する実施者として、ある営利を目的とした企業ではなく、昨今はCGM(Consumer Generated Media)の流行により利用者側で行われることが増えている。その代表例が「YouTube」である。「Youtube」は、インターネット上の無料の動画共有サービスであるが、メタデータとして、タグと呼ばれる動画コンテンツを分類するキーワードを、登録者だけでなく、誰でも自由に付加できるところが特徴的である。本節では「YouTube」の後を追って日本で開始された「ニコニコ動画」「テレビブログ」について説明する。

### 2.3.2 「ニコニコ動画」

本節では、動画の構造化の市場適用例として、「ニコニコ動画」について説明する。

#### (1) 本サービスの概要

2007年1月15日から株式会社ニワンゴが提供している動画配信関連サービス。2008年1月21日の時点で、一般会員登録者数が500万人を越え、有料会員も17万4000人を突破しているといわれる。動画を直接アップロードするサービスそのものではなく、他の動画投稿サイトにアップロードされた動画を独自のコメント表示・投稿機能を通して観覧させるというものである。

#### (2) 本サービスの特徴

本サービスの特徴は、動画の時間軸を意識したコメントの投稿・再生である。動画再生中にコメントボタンを押すことでコメントが投稿順に記録され、投稿したタイミングから3秒間動画上に表示される。コメント投稿そのものに時間差があっても、動画内の時間軸においては常に書き込まれた時と同じタイミングで表示される。その結果、閲覧者はチャットや掲示板のような時系列とは異なる擬似的な時間共有を体感することができる。コメントに時間の概念を導入したことで、ニコニコ動画は従来の動画投稿には無い「利用者同士の一体感」を醸成することに成功している。

また、コマンド機能を併用することで、すべての動画に対して文字色や文字の大きさ・表示方法などを数種類から選び、その動画の好きな位置に自由にコメントを行うことができることも特徴である。

なお、各動画には説明文のほか、タグと呼ばれる動画の内容を指し示す単語を10個まで登録するこ

とができる。動画投稿者だけでなく閲覧者も自由にタグを登録することができる。本来は検索機能として用いられるタグだが、このことを利用して、動画の内容に絡めたネタ的なタグ付けやニコニコ特有のタグ付け（「才能の無駄遣い」「名曲リンク」「再生数xxxx以下限定名作リンク」「もっと評価されるべき」など）をするなど、タグを動画投稿者と閲覧者、また閲覧者相互間のコミュニケーションツールとして用いられている事も特徴となっている。

検索機能では、動画につけられているタイトル、説明文、タグを元に、動画の検索ができる。検索結果は投稿日時及び再生回数のそれぞれで昇順、降順にソートすることが可能である。また、気まぐれ検索というニコニコ動画に登録されている動画を任意に抽出して検索することも出来るようになっている。

### 2.3.3 「テレビブログ」

本節では、動画の構造化の市場適用例として、「テレビブログ」について説明する。

#### (1) 本サービスの概要

株式会社メタキャストが提供するテレビ番組についての感想や意見を書くことに特化したブログポータル。テレビ番組ごとに専用ページを設けていて、番組ごとに内部ブログと、トラックバックを利用した外部ブログとを集め、一覧できるようになっている。

#### (2) 本サービスの特徴

本サービスは、関東地方とBS放送の全番組の専用ページを用意している。この専用ページがブログになっていて、外部にブログを持っていても、誰でもトラックバックできる仕組みになっている。

また、各番組の放映ログとして、「番ログ」を公開している。この「番ログ」は番組の主なシーンや出演者や告知、宛て先などの情報をテキスト化したものである。

#### (3) 動画関連ツール

株式会社メタキャストでは、以下のような動画関連のツール・サービスも一般公開している。

##### ①動画再生ソフト「TAGIRI」

PC上にある動画ファイルから自動的にサムネイルを作成する機能を持っている。15秒毎に作られたサムネイルが時系列順にリストとして並び、指定したポイントからすぐに再生ができる。テレビブログのサイトにある番組チャプターをダウンロードし、テレビ番組を録画した動画に設定することも可能である。

YouTubeやニコニコ動画、Google Videoなどのビデオ共有サイトから、簡単に動画をダウンロード

ドして保存できるツールバーも公開している。

## ②動画検索サービス「SAGURI」

TAGIRIツールバーの視聴履歴を使って、今、話題になっている「旬な動画」を中心に、複数の動画共有サイトに掲載されている動画を横断して一括検索できる動画検索サービス。最近の人気順、全期間の人気順、新着順で検索結果を並べることができる。

## 参考文献

- 1) ニコニコ動画：“ニコニコ動画ホームページ”，<http://www.nicovideo.jp/>，2008
- 2) WikiPedia：“「ニコニコ動画」Wikipedia”，  
<http://ja.wikipedia.org/wiki/%E3%83%8B%E3%82%B3%E3%83%8B%E3%82%B3%E5%8B%95%E7%94%BB>，2008
- 3) テレビブログ：“テレビブログホームページ”，<http://www.tvblog.jp/>，2008
- 4) 株式会社メタキャスト：“株式会社メタキャストホームページ”，<http://www.metacast.co.jp/>，2008

(橋本三奈子)

## 2.4 ISO/TC37/SC4における国際標準化

ISO/TC37/SC4では言語資源管理に関するさまざまな国際標準の策定が進められている。以下ではそのうちTDG3 (Thematic Domain Group 3; 意味内容記述のためのデータカテゴリ) における談話関係の標準的なデータカテゴリについて述べる。TDG3では談話関係や意味役割 (semantic role) 等の標準的なデータカテゴリを策定しているが、これらはISOの通常の国際標準よりも頻りに内容が変わると考えられるため、ISOの公式の国際標準ではない標準的なリポジトリとすることが想定されている。

下の表に談話関係の名称と定義と用例と出典を日本語と英語で列挙する。この表では、Sanders(1997)の理論に従い、談話関係を大きく4種類に分けている。つまり、談話関係全体を因果的 (causal) なものと付加的(additive)なものに分類し、その各々を肯定的 (positive) なものと否定的 (negative) なものに分類した。また、詳述 (elaboration) は付加的・肯定的な関係の一部である。

用例における`1`と`2`はそれを左端に含む[]の内部が当該の関係の第1項と第2項であることを示し、[]の外の部分は関係そのものを表わす。用例の間は改行で区切ってある。出典は各関係と何らかの意味で対応する関係を論じた文献を示す。ここで、`-`が関係の前に付いている場合はその逆関係を表わす。特に断らない限り、これらの対応関係の認定においては重要性と事実性(factuality)を捨象している。たとえば、RSTのPURPOSEという関係は第2項 (satellite) が非事実的 (non-factual) であるという制約を含むが、下記の「目的」(purpose) 関係はそのような制約を含まない。また、`RST88`はMann & Thompson (1988)、`RST01`はCarlson & Marcu (2001)、`RST`は両方を意味する。「市川」は市川(1990)である。

関係	定義	用例	出典
<b>因果的</b>			
因果・肯定(順接)			
順接	1は2の原因/証拠/動機/手段/可能要件/きっかけ/条件である。因果・肯定関係そのもの。	(「結果」～「条件推論」のすべての用例)	
derives	1 is a cause/enablement/motivation/evidence/trigger/means/condition of 2. The causal-positive relation.	[2 Tom passed the exam] because [1 he studied hard]. [1 Tom studied hard]. So [2 he passed the exam].	

因果	原因1が結果2を引き起こす。	[1 彼は試験に合格した]。かくて[2 彼の宿願は達せられた]。 [2 春子は知っていた]。なぜなら[1 本で読んだことがあった]から。	
Causes	1 is a cause of 2; 2 is a result of 1.	[2 Tom passed the exam] because [1 he studied hard]. [1 Tom studied hard]. So [2 he passed the exam]. [2 It' s raining] because [1 people are putting up umbrellas]. As [1 Tom passed the exam], [2 he must have studied hard]. [1 Here is a glass of wine]. [2 Drink it]. [2 Let' s have a break] because [1 we have half an hour].	The union of RESULT, CONSEQUENCE, -ENABLEMENT of RST and MOTIVATION of RST88.
可能化	1によって2が可能になる。「因果」の一種。	[1冷蔵庫にビールが入っている]から[2飲んで下さい]。	
Enables	1 makes 2 possible or permits 2. Kind of `causes' .	[1 Here is a glass of wine]. [2 Drink it]. [2 Let' s have a break] because [1 we have half an hour]. [2 I bought some beer] because [1 I had some extra money].	-ENABLEMENT of RST
動機付け	1が2を動機付ける。「因果」の一種。	[2腹が減っていた]ので[1レストランに入った]。	
Motivates	1 is a motivation of 2. Kind of `causes' .	[2 I bought some beer] because [1 was thirsty].	MOTIVATION of RST88 and - REASON of RST01
誘発	きっかけ1から2が生ずる、または1の出力が2。1は2の原因に関する具体的な情報を含まない。	[1 並木道を歩いて行った]。すると[2むこうからひとりの男が近づいてきた]。 [1 試食してみ]たら[2 なかなかいける]。 [1 実験]の結果、[2 濃度が2倍になった]。	市川
triggers	2 occurs with 1 as a trigger, or 2 is an output of 1. 1 contains no concrete information about 2' s cause.	[1 I was walking along the avenue]. Then [2 a man approached me]. [2 The density doubled] as a result of [1 the experiment].	

推論	根拠1から結論2が推論される。	[1山口君も来たということだ]。とすると[2この話を知らぬはずはない]。	
Inference	2 is inferred from 1.	[2 It' s raining] because [1 people are putting up umbrellas]. As [1 Tom passed the exam], [2 he must have studied hard]. [1 Here is a glass of wine]. [2 Drink it]. [2 Let' s have a break] because [1 we have half an hour].	JUSTIFY of RST88, CONCLUSION of RST01, and - EVIDENCE of RST.
演繹	1から2が演繹される。	[1雨が降ってきた。] [2傘が要る。]	
deduction	1 is inferred to cause/enable/justify /motivate 1. Kind of `inference' .	[1 A big meteorite hit the earth.] [2 Most dinosaurs would be dead then.]	
仮説		[1傘をさしている人がいる。] [2雨が降っているのだろう。]	
abduction	2 is assumed to cause/enable/justify /motivate 1. Kind of `inference' .	[1 Most dinosaurs died about 65,000,000 years ago.] [2 Maybe a big meteorite hit the earth.]	Subsumes - EXPLANATION - ARGUMENTATIVE of RST01
評価	1は2と評価される。	[1「まかめ種は生えぬ」という]。[2そのとおりだと思う]。	
Evaluation	2 is an evaluation of 1. Kind of `inference' .	[1 Tom passed the exam.] [2 He did very well.]	RST
解釈	1は2と解釈される。 「推論」の一種。		
Interpretation	1 is interpreted as 2 from a different perspective. Kind of `inference' .		RST
目的	手段1の目的が2である。	[2 五時に着きたい]。そのためには [1三時に家を出なくては]。 [2 合格する]ため[1 勉強した]。	
Purpose	1 is a means for the purpose of 2. Unlike in RST, 2 may be factual.	[1 Tom studied hard] in order to [2 pass the exam]. [2 Tom passed the exam] by [1 studying hard].	PURPOSE of RST and - MEANS of RST01
条件推論	もし1ならば2である。	[1太郎が女]なら[2美人だろう]	
conditional	If 1 then 2.	If [1 Tom comes here], [2 he' ll be surprised].	-CONDITION of RST88. The union of

			- CONDITION and - HYPOTHETICAL of RST01.
因果・否定(逆接)			
相反	1にもかかわらず2。 1と2が両立しているがそれはありそうにない または望ましくない。 対称。	[1わたしは全力をつくした]。それなのに[2結果は思わしくなかった]。 [1天気予報では晴れるはずだった]。ところが[2 雨になった]。 [1もう一度調べる必要がある]。[2結果はわからない]が。	
Conflict	1 and 2 cooccur, which is unlikely or undesirable. Symmetric.	[1 Tom studied hard] but [2 he failed in the exam]. [1 Tom failed in the exam] despite [2 his hard work].	CONSESSION of RST
非条件推論	1かどうかによらず2。	[1 天気が良い]かどうかによらず[2 行こう]。	
Unconditional	2 whether or not 1.	[2 I' ll go] whether or not [1 it rains].	RST88
譲歩	たとえ2であっても1である。	[1 天気が悪く]ても[2 行く]。	市川
compromise	2 whether or not 1, where 1 may prevent 2.	[2 I' ll go] even if [1 it rains].	
反条件推論	1でなければ2。		
Otherwise	If not 1 then 2.	[1 Go away.] Otherwise [2 I' ll leave].	RST
付加			
付加・肯定			
付加	1と2が共通の主張を強める。 対称。	[1かれはフランス語が得意だ]。また[2ドイツ語もよくできる]。 [1空はよく晴れていた]。[2涼しい風が吹いていた]。 [1三郎はひどく疲れていた]。そのうえ、[2少し熱があった]。	
Addition	Both 1 and 2 hold and jointly support something in common. Symmetric.	[1 The weather was fine], and [2 the wind was cool]. [1 Tom was tired]. Also [2 he was feverish].	Union of CONJUNCTION and ADDITION of 市川
補充		[1第一回の研究会を十六日に開きます]。なお、[2次回は来月の予定です]。	
Supplement	2 provides information relevant	[1 The first meeting shall be held on the 16 <sup>th</sup> .] [2 The next one shall	ELABORATION-ADDITIONAL of

	to 1	be in next month].	RST01.
背景	2が1の背景説明である。	(第1項を受話者が理解するための文脈を第2項が与える。「因果」や「推論」ではない。) [2太郎は中央線で新宿まで通勤している]。 [1今日は中野で下車した]。 [2最近本を三冊読んだ]。 [2その中に『友情について』というのがある]。 [2夏休みには旅行したいと思います]。 [1いっしょにいきませんか]。	
Background	2 provides background information of 1.	[2 My hobby is fishing.] [1 I went to the pond to catch some carps.]	RST
内容	1は発言や思考であり、2はその内容である。	[1その人は言った]。 [2「いつかお目にかかりましたね。」] [2結婚したい]という[1願望] [1太郎は[2「面白い」]と思った]。	
Content	1 is an utterance, a thought, or the like, and 2 is its content.	[1 I think] that [2 she is wrong]. [1 the idea [that [2 the earth revolves around the sun]]] [1 a desire [2 to get married]]	-ATtribution of RST01.
比例	2の程度に応じて1の程度が大きい。	[2大きい]ほど[1良い]。	
Proportion	The degree of 1 is in accordance with that of 2.	[2 the sooner], [1 the better].	RST01
類似	1と2が似ている。対称。	健は[2奈緒美]ぐらい[1大きな家に住んでいる]。 [2蝶]のように[1舞う]。	
Similar	1 and 2 are similar to each other. Symmetric.	[1 dance] like [2 a butterfly] [1 She cooks a turkey] as [2 her mother did].	ANALOGY of RST01
様態	2の様態で1。	[2寝]て[1暮らす]	
manner	1 in a manner similar to that of 2. (Same as `similar' ?)	[1 dance] like [2 a butterfly] [1 She cooks a turkey] as [2 her mother did]. [1 Do it] exactly as [2 I said]. ( `[1 Tom loves Mary] [2 very much].` Is not a case of `manner` . Rather, `[2 Tom loves Mary] [1 very much].` Is a case of `object` .)	RST01
詳述 (elaboration)			
詳細	2は1の詳細で、1は2の	[2この文章は、しくみが粗雑なうえ	

	要約。	に、用語も不適切だ。要するに、[1 悪文だ]。	
Detail	2 is details of 1; 1 summarizes 2.	[1 I bought some beer.] [2 I paid US\$10 and got four cans of Guinness.]	-SUMMARY of RST
具体論	2は1の具体的側面。	[1母は病院で働いています]。[2看護師です]。 [1 僕はりんごがきらいだ]。むしろ [2 みかんが好きだ]。	
Specific	2 is a specific aspect of 1.	[1 The projects are big.] [2 They can be \$1 billion.] [1 I hate bananas.] Rather [I'd like ananas].	GENERAL-SPECIFIC of RST01
部分	2は1の部分。		
Part	2 is a part or constituent of 1	[1 They all escaped.] [2 Some went to the seashore.]	Subsumes - PART-WHOLE of RST01
要素			
element	2 is a member of set 1.	[1 They all escaped.] [2 Mary went to the seashore.]	SET-MEMBER of RST
例	2は1の例。		
Example	2 is an example of 1.	[1 I like Japanese foods.] For example, [2 I love sushi].	RST01
抽出		[1すばらしいながめだ]。とりわけ、 [2林と丘はまるで絵のようだ]。	市川
extraction	2 is a particular example of 1. Kind of 'example' .	[1 I like Japanese foods.] In particular, [2 I love sushi].	
最小限	2は1の最小限の例	[1良い成績を取りたい。]少なくとも [2数学は落としたい]。	市川
minimum	2 is a minimal possibility of 1.	[1 We'll buy you some food]. [2 You'll have some sandwiches] at least.	
ステップ	2は過程1のステップ	[1水は循環している]。[2雨や雪として空から地上に降り、蒸発して空に戻る]。	
Step	2 is a step of a process 1.	[1 Water is circulating.] [2 It falls on the earth as rain or snow, and evaporates back to the sky.]	PROCESS-STEP of RST01
定義	2は1の定義	[1騙す]とは[2嘘を信じさせる]ことである。	
Definition	A definition of 1 is 2.	[1 To lie] is [2 to tell a false].	DEFINITION of RST01

換言	1と2は同程度に詳しい 言い換え。 対称。	[1私は奈緒美の夫だ。] つまり[2奈 緒美は私の妻だ]。	
Restatement	1 and 2 paraphrase each other with similar fineness. Symmetric.	[1 I' m married with Susan.] That is, [2 she is my wife].	RST
付加・否定			
選言	1または2。 対称。	[1思いきって言おうか]。それとも、 [2やはり黙っていようか]。	
Disjunction	1 or 2. Symmetric.	[1 Publish] or [2 perish].	RST01
対照	1と2が対照をなす。1と 2が両立することは大 いにありうるまたは両 立しても差し支えない が、両者は対照をなす。 両者を「のに」で結ぶ ことはできない。 対称。	[1値段が高い]が[2品質が良い]。 [1六人がその案に賛成だと言った]。 一方、[2それに反対する人は三人に 過ぎなかった]。	
Contrast	1 and 2 are in contrast to each other. Their cooccurrence is neither unlikely nor undesirable, so that they cannot be connected with 'despite' . Symmetric.	[1 His hair is white] but [2 his eyes are black].	RST (ANTITHESIS of RST has no equivalent here.)
相違	1と2が似ていない。 対称。		
Dissimilar	1 and 2 are dissimilar to each other. Symmetric.	[1 Tom is rich] unlike [2 Sue].	
優等比較	2の程度よりも1の程度 の方が大きい。	健は[2奈緒美]より[1大きな家に住 んでいる]。(「奈緒美」は「奈緒美 が住んでいる家」の意味。)	
moreThan	The degree of 1 is greater than that of 2.	[1 Tom swims faster] than [2 Sue walks].	Kind of COMPARISON of RST01 Subsumes PREFERENCE of RST01

ましてや		[2この問題は兄にも解けない]。まして [1ぼくにわかるはずがない]。	市川
letAlone	2 is more granted than 1.	[1 He can' t read], let alone [2 write].	
制約	2は1の制約。	[1わたしはサッカーが好きです]。ただし [2見るのが好きなんです]。	市川
constraint	2 constrains 1.	[1 I can come with you] though [2 I must go home by eight].	
例外	2は1の例外である。	[1このクラスは総じて優秀だ]。ただし [2太郎は出来が悪い]。	市川
exception	2 is an exception of 1.	[1 Boys like girls], except that [Tom hates Mary].	

### 参考文献

Lynn Carlson and Daniel Marcu (2001) Discourse Tagging Reference Manual.

<http://www.isi.edu/~marcu/discourse/tagging-ref-manual.pdf>

市川 孝(1990) 文章論概説. 教育出版.

William C. Mann and S. A. Thompson (1987) Rhetorical Structure Theory: Description and Construction of Text Structures. In G. Kempen (ed.) Natural Language Generation: New Results in Artificial Intelligence, Psychology, and Linguistics, 85-96, Martinus Nijhoff.

Ted Sanders (1997) Semantic and Pragmatic Sources of Coherence: On the Categorization of Coherence Relations in Context. Discourse Processes, 24, 119-147.

(橋田浩一)

## 2.5 利活用技術

### 2.5.1 E-learningと評価 — E-Learningの実践例の紹介 —

本節では、2007年9月に行われた第23回日本教育工学会全国大会から本委員会と関連がありそうな話題をいくつかピックアップし、教育工学における最近のe-Learningの現状を見ていく。

特に、e-Learningを実際に授業に取り入れ、活用している事例を紹介し、その利点、問題点等を概観する。

#### 国際医療福祉大学 遠隔授業システムからインターネットによる授業配信への展開 pp. 345

背景：キャンパスが分散（東京、栃木）。社会人学生が多い。

方式・特徴：ビデオ・オン・デマンド(VOD)とインターネット中継による授業配信。VODはWMV形式で提供。補助教材の配布(PDF)。授業後アンケートによる受講確認。蓄積型51講義と中継型3講義の提供。登録者は学生33名、学外10名、教員17名。オープンソースの積極利用による低コスト化。

評価、問題点：教育システムとしては、実用域にあるが、プログラムが少ない、VODでは、教員・受講者のコミュニケーションが難しいなどの問題が明らかになった。補完授業としては有効か。

#### 青山学院大学 オンライン学習支援者と学習者の意識差に関する事例研究 pp. 341

背景：eLearning学習においては支援者の存在が重要。

方式・特徴：学習支援者と学習者がそれぞれ必要と考えるスキルについてずれがあるかどうか調査。Eラーニング専門家育成プログラム受講生（89名）および支援者（16名）を対象にアンケート調査を実施した。

評価・問題点：受講者は支援者を相談相手、話し相手として見る傾向があり、学習上の指導はあまり重視していないということが分かった。一方、支援者は学習指導を重視する傾向が強かった。

#### 梅花女子大学 「学び舎広場」の創造 pp. 337

背景：教材の共有化が動機。

方式・特徴：本学開発のテストオーサリングツール「楽々てすと君」を他大学、高校、個人に解放、当時に関連情報を共有化するコミュニティサイト「学び舎広場」の開設。テスト問題と解答を共有化。

評価・問題点：高校教育のEラーニング化推進に寄与。レメディアル教育への活用も期待される。高

校側に負担（デジタル化など）が発生。「てすと君」の改良にも有用と考えられる。

明治大学 eラーニング運営のための情報可視化手法の提案:明治大学ユビキタスカレッジ構想の事例  
から pp. 349

背景：「明治大学ユビキタスカレッジ」構想の一貫。学習者支援環境の高度化が目標。

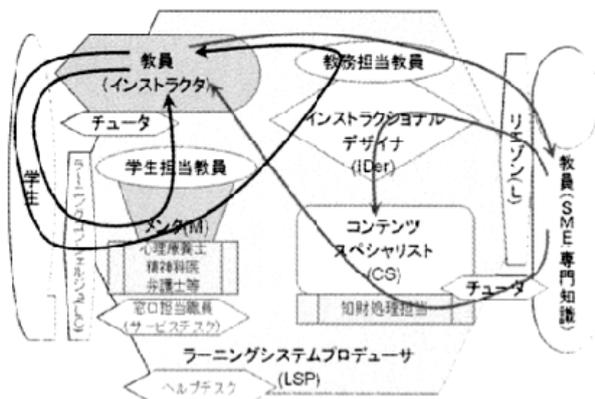


図1 eラーニング専門家チームが支援する組織

方法・特徴：教員の個人芸であった教授法を、左記の図のように、知財、法律、心理の専門家を含む大掛かりなチーム活動に再編し、教育効果を向上させる。2007年度後期より実施。チームメンバーの役割を明確化、相互監視する。

評価・問題点：評価は、学習効果、情報流通、学生満足度の観点から、今後実施する予定。

岐阜大学 大学における教育支援システムの統合化 pp.351

背景：総合情報メディアセンター設置を契機として、学生向けの支援システムを構築。

方式・特徴：学務、保健、教養、図書の情報システムに集約。履修登録機能の実現。

シングルサインオン機能を導入し、図書館ポータル、電子メール、他学内情報サービスにシームレスにアクセス可能。アンケート機能の実現。

評価、問題点：非正規学生、非常勤教職員へのサービスが未整備。施設情報の統合がまだ。

金沢大学 e-Learningの全学展開を考慮した教育システムの構築 pp. 531

背景：金沢大学では2006年度から新生に携帯型PCの必携化が始まった。「情報処理基礎」の授業開設を契機として全学規模のE-Learningシステムの導入。

方式・特徴：いつでも、どこでも、学習できる環境構築を目指し、学内の主要な施設、講義室、図書館、ロビー、生協などに無線LANのアクセスポイントを設置。現在66箇所に設置。「情報処理基礎」

は2教室を併用、TV会議システムを使って行う。他、サポート窓口を設けている。

評価、問題点：2006年度と2007年度に使い易さに関するアンケート調査を学生に対して実施。79%が高評価。また、システムについては、53%が自宅でも利用したと回答している。

### 京都大学 Webを利用した公開授業システムの実用化に向けて pp. 523

背景：京都大学では大学教育改善活動の一環として、「公開講座・討論会」を実施してきたが、参加者が同じ時間に特定の場所に居なければならず、障害となっていた。この問題に対応するためウェブ版の公開講座を構築した。



図1 ビデオ視聴画面

方式・特徴：左の図にあるように、授業映像は2つのアングルで提示。授業関連資料はダウンロード可能な電子ファイルとして用意。視聴者は、BBSで討論に参加。

評価・問題点：議論が盛り上がり、不発。参加者の帰属意識の形成が課題。

### 新潟大学 教授・学習活動を総合的に支援する統合型学務情報システム pp. 769

背景：現在、大学が取り組んでいる新たな学士課程教育を目指した教育改革をバックアップすることが目的。

方式・特徴：同一認証によるウェブアクセシビリティの向上。容易にカスタマイズ可能。データベースの一体化。ポータルシステムと学生カルテシステムが主機能。ポータルシステムは授業関連情報の集約を実現、カルテシステムは、個々の学生の成績、履修関連を一元管理する。

評価・問題点：個人情報扱いが課題。カルテシステムの活用法については、まだ手探り。

### 佛教大学 社会的要請に応じるe-Learningを用いた高等教育 pp. 119

背景：大学は、高等教育の授業料の高騰をいかに押さえるか、幅広い年齢層の学生に対応した授業を

どう行かうかという問題に直面している。E-Learningに問題打開の可能性。

方式・特徴：通信教育の講義を一部 E-Learning化。E-Learningと対面授業の併用。E-Learningは主にビデオ講義、ウェブ上でのテスト、電子化された講義資料の提供、掲示板設置より成る。対面授業はスクーリングにより実施。

評価・問題点：受講者数は、265名。アンケート調査の結果、良好な評価を得た。サポート体制を含めた大学全体の組織的な取り組みが必要。

#### 岐阜大学教育学部 働きながら学ぶ教師を対象した遠隔講義の授業設計 pp. 115

背景：現代GPの研究をベースとして昼間通学が難しい現場教師向けのインターネット大学院を開講。

方式・特徴：対面講義とインターネット講義を併用（いわゆる、blended learning）。受講者の負担軽減のため、対面講義は少なめしている。システムは、岐阜大に導入している学習管理システムAIMS-Gifuを活用。カリキュラムは以下参照。専門科目、課題研究以外はインターネットで受講。

		修士1年		修士2年	
		前学期	後学期	前学期	後学期
専門科目 (8科目)	同期 非同期	2科目 週1日 18:10-21:10	2科目 週1日 18:10-21:10	2科目 週1日 18:10-21:10	2科目 週1日 18:10-21:10
	対面	1科目 (夏期休暇)		1科目 (夏期休暇)	
課題研究 (3科目)	同期 対面		1科目	1科目	1科目
自由選択科目 (4科目)	非同期	← 2年間のうちに4科目 →			
共通科目 (2科目)	非同期	← 2年間のうちに4科目 →			

図1 eeeの2年間の受講イメージ

評価・問題点：受講生、教師のコミュニケーションで若干問題（回答がタイムリーではないなど）。受講生に対する、学習相談を強化していく必要がある。

#### 岩手大学 教育支援システム「In Assistant (アイアシスタント)」の全学導入 pp. 123

背景：文科省特別教育研究経費によるプロジェクトが発端。授業改善と教室外学習支援を目的としてシステムを構築。

方式・特徴：それまでばらばらに運用されていたウェブシラバスを統一。授業関連情報の提供と教師、

受講者間のコミュニケーション、学習環境、例えば、掲示板、WIKI、ファイル共有など、の提供を目的とする。ポータル画面の個人適用を実現した。

評価・問題点：サポート体制は未整備。420名の教員に対して、現在1人。シラバス作成については一定の成果。「従来のウェブシラバスシステムが新しくなっただけ」という声もあり、周知が不徹底。全教員が参加していない。

(野本忠司)

## 2.5.2 対話型学習システムを構築するツールTuTalkについて

本稿では、教育用対話システムを簡単に作成できる TuTalk (<http://www.learnlab.org/technologies/tutalk/index.php/>) というツールを紹介する。これは米国の Pittsburgh Science of Learning Center で開発が進められているもので、2005年の夏に最初のプロトタイプが完成し、10名の学習者によって評価実験が行われた。それ以降、順調に開発が進行し、2008年夏にはツールを実体験するサマースクールが計画されている。

TuTalkを使って構築するシステムでは、学習者がシステムと自然言語で対話し、その対話を通して学習が進行する。対話管理部の作成に際しては、プログラミング能力はほとんど必要としない。

ツールはモジュールベースで構成され、各モジュールはエージェントとして動き、特定のサービスを提供するとともに、他のモジュールへ要求を出すようになっている。サービスの授受管理は調整モジュール (Coordinator) が担当し、他のモジュールとしては入力、出力、ログ管理、入力解析、出力生成、学習者モデル、対話管理、対話履歴管理等がある。以下に、重要なモジュールならびに対話設計方針について考える。

### 入力解析モジュール

学習者は自由に自然言語文を入力してよいことになっているものの、現在の解析モジュールは入力に対して、最小距離法を用いての最も関連の深いコンセプトへマッチングするだけとなっている。自然言語文の入力を許しているといいながら、その解析がこのような単純な機構になっているのはもったいない。改良の余地が多いと言える。

### 対話管理モジュール

TuTalkの対話管理モジュールは、システムからの質問と学習者からの回答をノードとしたようなスタック付きの有限状態ネットワークとなっている。対話管理が有限状態ネットワークでできるということは、タスク自体の構成が単純で、システム主導の対話に限られることが多いが、TuTalkでは次のようないくつかの付加的な仕組みが備わっている。すなわち、学習者が部分的にだが主導権を握るこ

とができ、また、学習者の回答が不正解の場合に副ネットワークに移動できるようになっている。

### コンセプトの設計

有限状態ネットワークのノードは対話のコンセプトに相当する。これは単に自然言語の句または単語列の集合である。

Skip\_appetizer

```
[phrase=minimal, I not appetizer]
[lang=fr, Je n'ai pas le temps pour un amuse-gueule]
[processing-type=extended, principle1 principle2]
```

上の例で、skip\_appetizerはコンセプトIDであるが、“I”, “not”, “appetizer” という3語を含む入力（例えば “I would not like any appetizer”）にマッチすることを表している。デフォルト言語以外の場合には、2行目のように言語名を指定して記述することができる。さらに、principle1といった他の処理へのポインタを書いて、複雑な処理を行うこともできる。

### シナリオの設計

シナリオ (recipe) は基本的には下の例のようにステップごとに子シナリオへのポインタで記述したり、返答の種類に応じた対処を[]で記述したりする。例中の unknown というのは、どの返答パターンにも当てはまらない場合の処置である。例の最下行のように副ゴールを設定し、まるでサブルーチンを呼ぶような記述も可能となっている。

Recipe : Select-appetizer

```
Step: enthuse_about_appetizers
Step: ask_share_appetizer
    [agree_to_share_appetizer]
    [skip_appetizer: abort, ask-soup]
    [unknown: abort, loose-temper]
Step: subgoal: agree-on-appetizer
```

### 参考文献

- 1) P. Jordan, B. Hall, M. Ringenberg, Y. Cui and C.P. Ros “Tools for Authoring a Dialogue Agent that Participates in Learning Studies”, In Proceedings of Artificial Intelligence in Education,

AIED 2007, pp. 43-50 (2007).

- 2) Kersey, Cynthia, Di Eugenio, Barbara, Jordan, Pamela and Katz, Sandra “Collaboration in Peer Learning Dialogues”, Presented at DECALOG – The 2007 Workshop on the Semantics and Pragmatics of Dialogue (2007).
- 3) Jordan, Pamela, Ringenber, Michael and Hall, Brian “Rapidly Developing Dialogue Systems that Support Learning Studies”, In Proceedings of ITS06 Workshop on Teaching with Robots, Agents, and NLP (2006).

(齋藤博昭)

## 2.6 作業研究

### 2.6.1 概要

本分科会では昨年度より、セマンティックオーサリング技術を用いたビデオコンテンツの構造化方式について検討を重ねている[1]。またプロトタイプ実装も継続している[2][3]。セマンティックオーサリングとは、文や文間の関係をラベルとする有向グラフとしてコンテンツを生成する行為をいい、また生成されたコンテンツを知的コンテンツと呼ぶ[4]。セマンティックオーサリングでは、伝えたい内容の論理構造を保存するので、人間と計算機の双方にとってコンテンツが分かりやすい。

セマンティックオーサリング技術の応用として、これまでに情報の要約や検索、アノテーションといった分野を中心に研究が行なわれてきた。これらはテキストコンテンツを対象としているものだったが、動画のショットをメタデータとフレーム画像により表現することで、対象をマルチメディアコンテンツに拡張する。これをセマンティックビデオオーサリングと呼ぶ。本年度も当技術を基盤としたビデオデータの構造化方式や応用技術に関する作業研究を行ったので、報告する。

### 2.6.2 データ構造化

国立情報学研究所(NII)より提供されている評価用ビデオコンテンツ[5]の中から日本伝統工芸のドキュメンタリー映像「音色」を用いてデータ構造化方式についての議論を続けている。今年度ショット情報に関するスキーマを再度策定した。スキーマを表2.6.2-1に示す。MPEG-7やDublin Coreなど既に広く流通しているメタデータスキーマ極力利用することにした。

映像の構造化方式についても今年度再検討した。昨年度はショットを単位として、他のショットとの談話構造を考慮してグラフ構造を構築する方式を採用してした。談話関係はISO/TC37/SC4/TDG3で策定中の関係セット（約50種類）を用いている。

従来のグラフ構造オーサリングでは自由度が大きいため、編集を続けるに従って大域的な構造を把握しづらくなり、また関係を示すリンク数も増えるにつれ接続関係もわかりづらくなるという点が指摘された。そこで、ショット群を階層的にグルーピングして木構造を構成し、表示位置を固定した上で、必要に応じて談話関係を付加する方式とした。我々の分科会で使用しているオーサリングツールであるセマンティックエディタにも木構造の専用作成モードを搭載した。図2.6.2-1にセマンティックエディタで木構造化した例を示す。左側のウィンドに表示されているのが木構造データである。ショットを構成する代表的なフレーム画像を木の深さに応じて自動的に配置する。葉以外のノードは、グループを構成する体表ノードのフレーム画像を表示している。葉のノードがショットの実体であり、開始時刻、終了時刻、ショットの説明文などのメタ情報が付加されている。セマンティックエディタの特徴は単なるコンテンツエディタにとどまらず、それに関するオントロジーやスキーマも同様の操

作をもって同時に定義できることである。右側に列挙されている内部ウィンド上がそれで、上から談話構造、MPEG-7、Dublin Coreに関する語彙を表現している。これにより属性の可否や属性値の制約に関する支援を受けながらコンテンツを生成できる。

表2.6.2-1 策定したビデオコンテンツのスキーマ

属性	意味	個数	値
sa:uri	自身のURI	1	URI
rdf:type	型	1	mp7:Segmentの下位クラス
dc:title	タイトル	0..1	文字列
dc:description	説明	0..*	文字列
sa:text	テキストコンテンツ	0..*	文字列
mm:media	映像等のURI	0..1	URI
mm:begin	mm:media で指定される映像・音声の中の開始時刻	0..1	時刻(ISO8601形式) 又はmsec
mm:end	mm:media で指定される映像・音声の中の終了時刻	0..1	時刻(ISO8601形式)又はmsec
mm:channel	チャンネル、トラックの識別番号	0..1	整数
sa:part	部分(begin、end、channelと共存しない)	0..*	mp7:Segment型リソース
sa:head	グループの中で代表となるショット	0..*	mp7:Segment型リソース
sa:thumbnail	サムネイル画像	0..1	画像ファイル/mp7:Segment型リソース
dc:creator	制作者	0..*	文字列(vcard)
dc:rights	著作権	0..1	sa:Rights型リソース

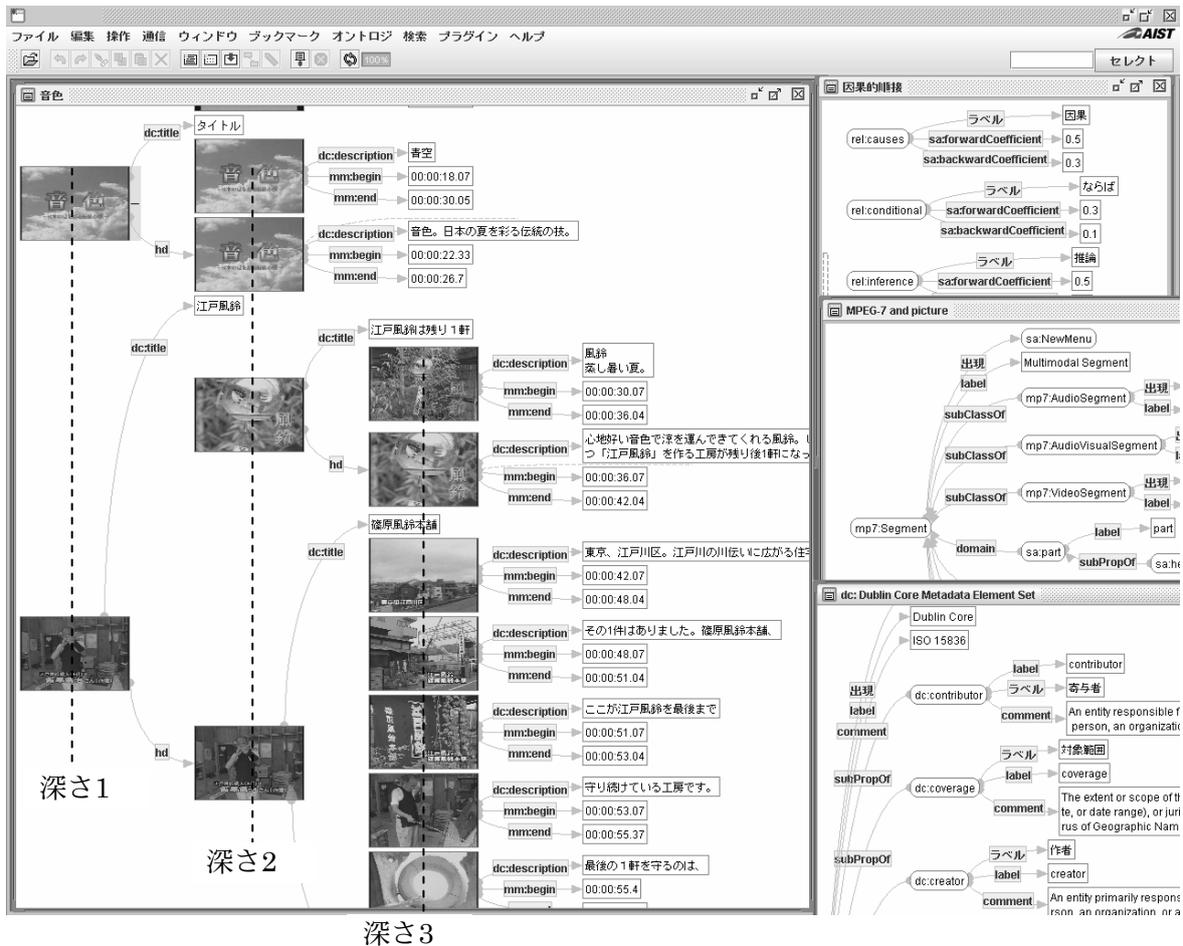


図2.6.2-1 セマンティックエディタを用いたビデオコンテンツの構造化例

### 2.6.3 プロトタイプ

本年度はコンテンツの高度利用方式の一例として、映像要約を実現するプロトタイプを実装した。スクリーンショットを図2.6.3-1に示す。



要約条件入力画面

要約再生画面

図2.6.3-1 プロトタイプシステムのスクリーンショット

本プロトタイプはブラウザ上で動作するWebアプリケーションである。図2.6.3-1の左側に提示したのが「要約条件入力画面」である。セマンティックエディタで作成したショット群をブラウザ上に木構造表示する。各ショットのサムネイル画像データや時刻範囲などの属性値の情報は、データベースより取得して所定の位置に表示される。要約に際して、重要とするショット（要約に含めたいショット）をチェックボックスにより選択する。入力キーワードとのマッチングによる重要ショット群の自動設定も可能である。

次に要約手法について略説する。設定された重要ショット群と要約の時間長から要約対象部分グラフを抽出し、ショットのシーケンスを決定する。各ショットに対応するノードの重要度は、活性拡散を用いて算定する。活性拡散の解説は本稿では割愛する。詳細については文献[6]を参考にされたい。図2.6.3-1の右側に提示しているのが「要約再生画面」である。再生順にショットを列挙して表示する。再生中のショットに対応する部分がハイライトされて遷移していく。

#### 2.6.4 まとめ

大量に流通・蓄積されるビデオコンテンツを効果的に活用するため、セマンティックビデオオーサリングによる映像コンテンツの構造化方式を検討した。また利用者の要求に基づいて要約ビデオを生成するWebアプリケーションを実装した。

今後は、要約アルゴリズムの再検討や、本件で扱った構造だけでなく発話内容まで考慮した重要性の評価を行なうことにより、さらに高精度な要約が実現されるであろう。また利用者の嗜好や閲覧履歴に応じた要約生成の仕組みについても検討を進める予定である。

#### [引用文献]

- [1] 07-情-5 「知識情報処理技術に関する調査研究報告書」, (社) 電子情報技術産業協会, p39-43 (2007年3月)
- [2] Hasida, K.: Distributed Semantic Authoring as Foundation of Semantic Computing ,in Notes on From Semantic Web to Semantic World Workshop conjoint with JSAI2003,(2003)
- [3] 伊藤一成, 藤原司, 橋田浩一: セマンティックビデオオーサリングによるニュース動画群からのダイジェスト生成, 人工知能学会 第74回知識ベースシステム研究会 SIG-KBS-A601,(2006)
- [4] 林志憲, 伊藤一成, Martin J. Dürst, 橋田浩一: 動画の対話的視聴システムSevinaの実装, マルチメディア, 分散, 協調とモバイル(DICOMO2007)シンポジウム論文集, pp. 605 – 610,(2007).
- [5] 馬場口 登, 栄藤 稔, 佐藤 真一, 安達 淳, 阿久津 明人, 有木 康雄, 越後 富夫, 柴田 正啓, 全炳東, 中村 裕一, 美濃 導彦, 松山 隆司: 映像処理評価用映像データベースについて, 電子情報通信学会技術研究報告, PRMU2002-30, (2002)

[6] J. R. Anderson: A Spreading activation theory of memory, *Journal of Verbal Learning and Verbal Behavior*, pp. 261–295, (1983)

(伊藤一成)

## 3. 言語資源

### 3.1 はじめに

言語資源分科会は、言語資源の研究開発利用に関わる問題点を調査することを目的としている。著作権に関する調査、言語資源カタログの整備、また言語資源に限らず、広く言語処理に関する情報提供の場としての、言語情報処理ポータル運営を行っている。これらの活動においては、言語資源協会（GSK）と協調しながら行った。

平成19年度もこれまでに引き続き、言語情報処理ポータル運営を行った。会議案内、製品ニュース、コラムの定期的な更新と共に、その他のコンテンツについても迅速な更新を行っている。当サイトは言語情報処理に関連する情報を網羅的・総合的に提供しており、当該分野および関連分野の研究者、開発者、学生などから好評を得ている。新しい試みとして、会議案内の欄で、メジャーなカンファレンスに対してのWikiページを作成し、皆で情報の更新や共有が可能な場を提供する試みを実施した。

また、言語資源カタログの整備を行った。この作業は情報通信研究機構や言語資源協会と協力しながら行った。情報通信研究機構では、名古屋大学と共同で、2007年から大規模言語資源メタデータベースSHACHIの構築を開始している。SHACHIは、日本・アジア諸国の言語資源をはじめ、世界中の言語資源の詳細なメタデータを収集し、得られた知見を基に体系的な蓄積を試みている。すでに2000件を超える言語資源が収録され、検索機能も装備している。SHACHIはすでに、<http://www.shachi.org> で公開しており、言語処理ポータルからもリンクが張られ、アクセス可能である。

言語資源についての幅広い知見を収集するため、有識者のヒアリングを行った。国立国語研究所の前川喜久雄氏にKOTONOHA計画における著作権処理について講演していただいた。日外アソシエーツの星俊雄氏に言語資源配布の実態について講演していただいた。このヒアリングをきっかけに、スポーツ報知新聞社のデータが研究利用可能となった。

以下では、これらの活動の詳細を述べる。また、最後に、平成20年度以降における活動テーマの可能性について述べる。

(執筆者 井佐原均)

## 3.2 活動報告

### 3.2.1 言語情報処理ポータル

言語情報処理ポータル ([http://nlp.kuee.kyoto-u.ac.jp/NLP\\_Portal/](http://nlp.kuee.kyoto-u.ac.jp/NLP_Portal/)) (図3.2.1-1) の運営を行った。言語情報処理ポータルは、言語情報処理に関するさまざまな情報を集約したポータルサイトで、2002年夏より公開している。担当委員が運営方針の策定と基本的なコンテンツ作成を行っている。

2008年2月時点で、次のような内容を掲載している。

- 会議案内
- 製品ニュース
- 新刊案内
- コラム
- 言語情報処理 用語集
- 講義資料リンク集
- 論文データベースリンク集
- 人材募集
- 言語資源カタログ
- プロジェクト・研究機関・学会等
- 世界の言語イニシアティブの紹介

上記のうち、会議案内、製品ニュース、コラムは定期的に更新を続けている。その他のコンテンツについても掲載すべき情報を見つけ次第迅速に対応している。このように当サイトは言語情報処理に関連する情報を網羅的・総合的に提供しており、当該分野および関連分野の研究者、開発者、学生などから好評を得ている。今年度は更新作業が中心であったが、安定して月平均2000程度のアクセスがあり、定期的にアクセスしている閲覧者が多いことが確認できる。

本年度は、会議案内の欄で、メジャーなカンファレンスに対してのWikiページを作成し、皆で情報の更新や共有が可能な場を提供する試みを実施した (図3.2.1-2, 図3.2.1-3)。Wikiページには、次のような内容が掲載されている。

- 会議の基本情報 (日時、場所、会議の公式ページへのリンク等)
- 追加情報 (締切日、ホテル情報等)
- 役立ち情報 (現地の治安状況、ビザの要否等)
- 疑問・質問エリア
- 関連リンク

今後、カンファレンスの詳細情報、現地情報、交通手段などの情報を入手する場として活用して頂きたい。

来年度のポータル活動としては、閲覧者の多いトップページの情報の更新作業を継続することと併せて、用語集で未定義の用語の説明を充実させる予定である。

The screenshot shows the homepage of the Natural Language Processing Portal Site. The header includes the site title in Japanese and English, and a language selector. The main content is organized into several sections:

- What's New:** A list of recent updates, including the addition of new terms to the glossary and the release of a new version of the portal.
- プログラム (Programs):** A list of upcoming conferences and workshops, such as the 9th Annual Tokyo Conference on Psycholinguistics and the 17th Pacific-Asia Conference on Knowledge Discovery and Data Mining.
- 用語集 (Glossary):** A section for defining terms related to natural language processing.
- お問い合わせ (Contact Us):** Information on how to reach the site administrators.

図3. 2. 1-1 言語情報処理ポータルサイト

This screenshot shows a list of conference links from the portal site. The links are:

- [SIGIR 08 \(The 31st Annual International ACM SIGIR Conference\)](#) (2008/7/20-24, Singapore, Singapore)
- [CoNLL-2008 \(Twelfth Conference on Computational Natural Language Learning\)](#) (2008/8/16-17, Manchester, England)
- [COLING 2008 \(The 22nd International Conference on Computational Linguistics\)](#) (2008/8/18-23, Manchester, England)
- [TSD 2008 \(11th International Conference on Text, Speech and Dialogue\)](#) (2008/9/8-12, Brno, Czech Republic)

A Wiki icon is visible on the right side of the list, indicating that the information is sourced from or linked to Wikipedia.

図3. 2. 1-2 Wikiページへのリンク

**COLING2008**  
[http://nlp.kuee.kyoto-u.ac.jp/NLP\\_Portal/pukiwiki/index.php?COLING2008](http://nlp.kuee.kyoto-u.ac.jp/NLP_Portal/pukiwiki/index.php?COLING2008)

[ トップ ] [ 編集 | 凍結 | 差分 | バックアップ | 添付 | リロード ] [ 新規 | 一覧 | 単語検索 | 最終更新 | ヘルプ ]

**最新の20件**

- 2007-12-21
  - COLING2008
  - ACL2008
  - FrontPage
  - MenuBar
  - RecentDeleted
- 2007-12-10
  - WWW2008
- 2007-12-08
  - WikiEngines
- 2007-11-17
  - IJCNLP-08
- 2007-10-30
  - LREC2008
- 2006-10-22
  - PukiWiki/1.4/Mannual
  - YukiWiki
  - InterWikiSandBox
  - PukiWiki/1.4/Mannual/Plugin/A-D
  - InterWiki
  - InterWikiName
  - Help
  - BracketName
  - PukiWiki/1.4/Mannual/Plugin/E-G
  - SandBox
  - PukiWiki/1.4/Mannual/Plugin/S-U

**COLING 2008 (The 22nd International Conference on Computational Linguistics) <sup>†</sup>**

**基本情報 <sup>†</sup>**

- 日時: 2008年8月18日 - 22日
- 場所: イギリス (マンチェスター)
- 会議のページ

**Important Dates <sup>†</sup>**

- 2008年3月30日、Submission deadline main conference
- 2008年5月30日、Notification of acceptance

**疑問・質問など <sup>†</sup>**

**関連リンク <sup>†</sup>**

- Wikipediaマンチェスター

Counter: 260, today: 2, yesterday: 2

Last-modified: 2007-12-21 (金) 11:02:19 (74d)

Link: FrontPage(74d)

図3. 2. 1-3 Wikiページ

### 3. 2. 2 言語資源カタログの整備

#### はじめに

近年、音声・言語の主要メディアに関する研究開発を目的に、電子化されたコーパス、辞書、シソーラスといった言語資源（音声・動画を含む）の構築が盛んに行われ、その重要性は広く認識されている。しかし、これまでに公開されている言語資源は、各開発機関において、個々の目的に応じてほぼ独立に構築されており、タグセットやフォーマットについても多種多様である。その一方で、言語資源の構築段階において、他の言語資源の仕様に準拠し、同じタグセットを参照するなど、実際には、個々の言語資源間に何らかの関連性があることも少なくない。

欧米では、Linguistic Data Consortium (LDC)、European Language Resources Association (ELRA) といった言語資源コンソーシアムが、主に欧米語の言語資源の収集、配布を行っている。また、Open Language Archives Community (OLAC) では、言語資源のメタデータの統一や、言語資源カタログの整備を担っている。しかし、カタログの情報は詳細ではなく、それらの言語資源が研究開発過程のどのようなフェーズで用いられたかといった情報や、他の資源との関係性などは示されていない。そのため、流通という面で十分機能しているとは言えない。国内においても、国立情報学

研究所（NII）や、言語資源協会（GSK）によって、音声・テキストコーパスを蓄積する活動が行われている。しかしながら、体系的に蓄積するには至っておらず、このような状況において、多様な目的を持ったユーザが、目的に合致する言語資源にたどり着くことは容易ではない。

そこで、情報通信研究機構（NICT）と名古屋大学は共同で、2007年から大規模言語資源メタデータベースSHACHIの構築を開始した。SHACHIは、日本・アジア諸国の言語資源をはじめ、世界中の言語資源の詳細なメタデータを収集し、得られた知見を基に体系的な蓄積を試みている。SHACHIのメタデータセットは、OLACのメタデータセットに新たに19項目を追加して拡張したものである。SHACHIに収録されている言語資源はすでに2000件を超える。また、検索機能も装備している。これにより、SHACHIには言語資源の流通拠点としての役割が期待される。SHACHIはすでに、<http://www.shachi.org> で公開しており、言語処理ポータルからもリンクが張られ、アクセス可能である。

本報告書では、SHACHIの目的、設計、メタデータの収集・拡張、及び、カタログ検索機能について述べる。

## SHACHIの目的

言語資源メタデータベースSHACHIを構築する目的は以下の5点にまとめられる。

- (1) **言語資源メタデータの蓄積**：既存の言語資源を有機的に結合し[1]、戦略的に言語資源を開発するためには、世界中の言語資源に関する情報が一個所にまとめられていることが重要である。そこで、大量の言語資源の詳細なメタデータを半自動で収集し、各言語資源の詳細なカタログを作成する。
- (2) **言語資源メタデータの体系化**：(1) によって収集された詳細なメタデータをもとに、言語資源のタイプを分類することにより、言語資源オントロジーの構築を試みる[2]。図1は、開発中の言語資源オントロジーの例である。現段階では、人手によって構築しているが、オントロジーを自動生成する手法の考案も試みている。

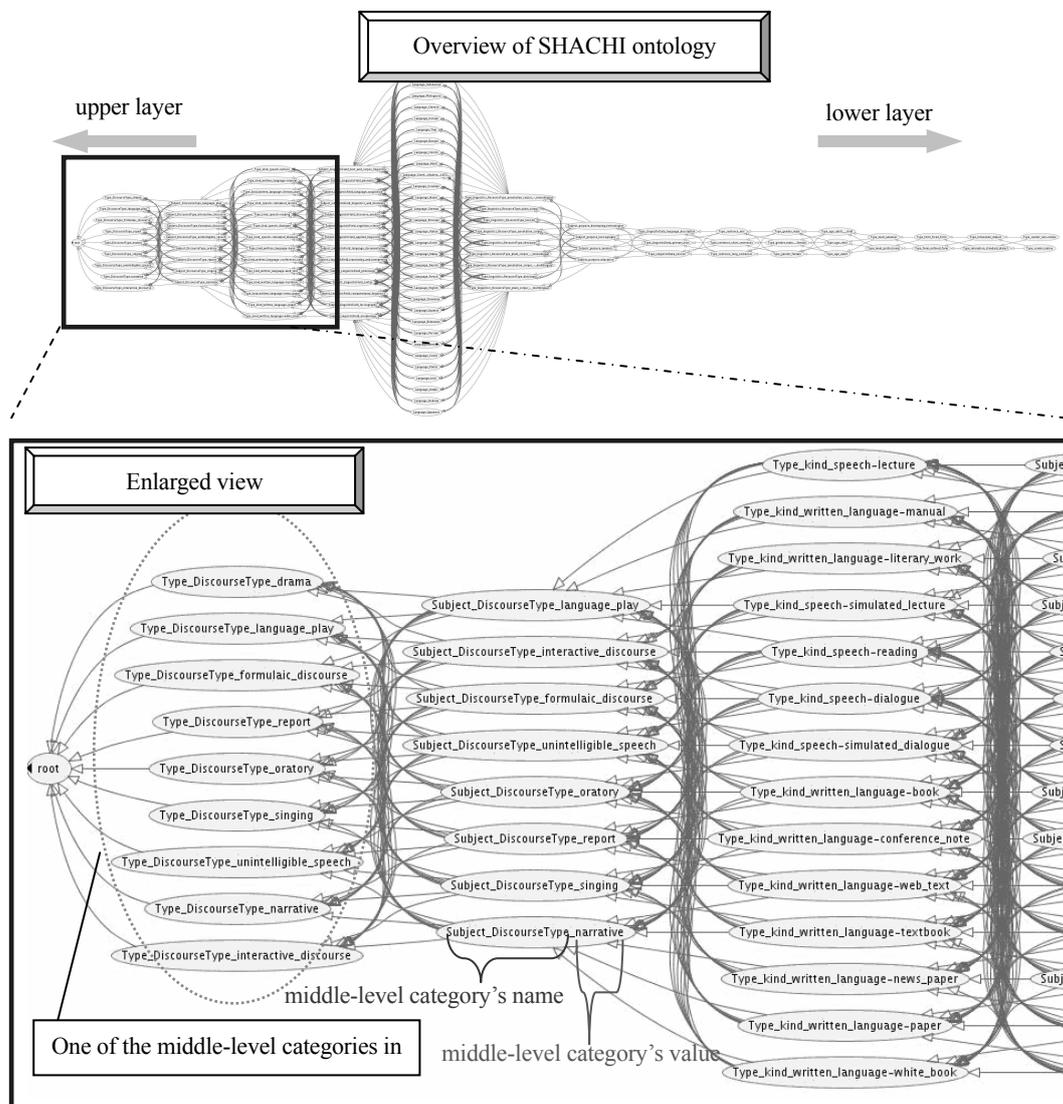


図3.2.2-1 言語資源オントロジーの構築（人手による構築の例）

(3) 言語資源間の関連付け：SHACHIはLDC、ELRAのカタログなどに比べ、詳細なメタデータを収集している点に特徴がある。これらの詳細なメタデータにより、個々の言語資源の詳細な特徴が記述できるようになり、さらには、言語資源間の関係性を明示化することができるようになることも期待される。図2は、SHACHIの検索画面の一部である。検索結果として表示された言語資源が準拠した資源（参照元）や、フォーマットが共通である資源を提示している。これらの言語資源間の関係性を統計的に調査することにより、世界標準レベルのタグセットや、データフォーマット、需要のある言語資源のタイプなどを策定できると考える。

(4) 言語資源の統計調査：SHACHIのサイトでは、収録されている言語資源メタデータに関する統計情報が閲覧できる。これらメタデータを統計的に分析することで、どのような資源が世界の

どこに存在するのかを把握したり、近年、公開されている言語資源の傾向を捉えたりすることができる。

- (5) 言語資源の流通促進：本メタデータデータベースに検索機能を整備し、ユーザのニーズに合致した言語資源へのアクセスを容易にすることにより、言語資源の有効利用や、効率的開発を支援する。

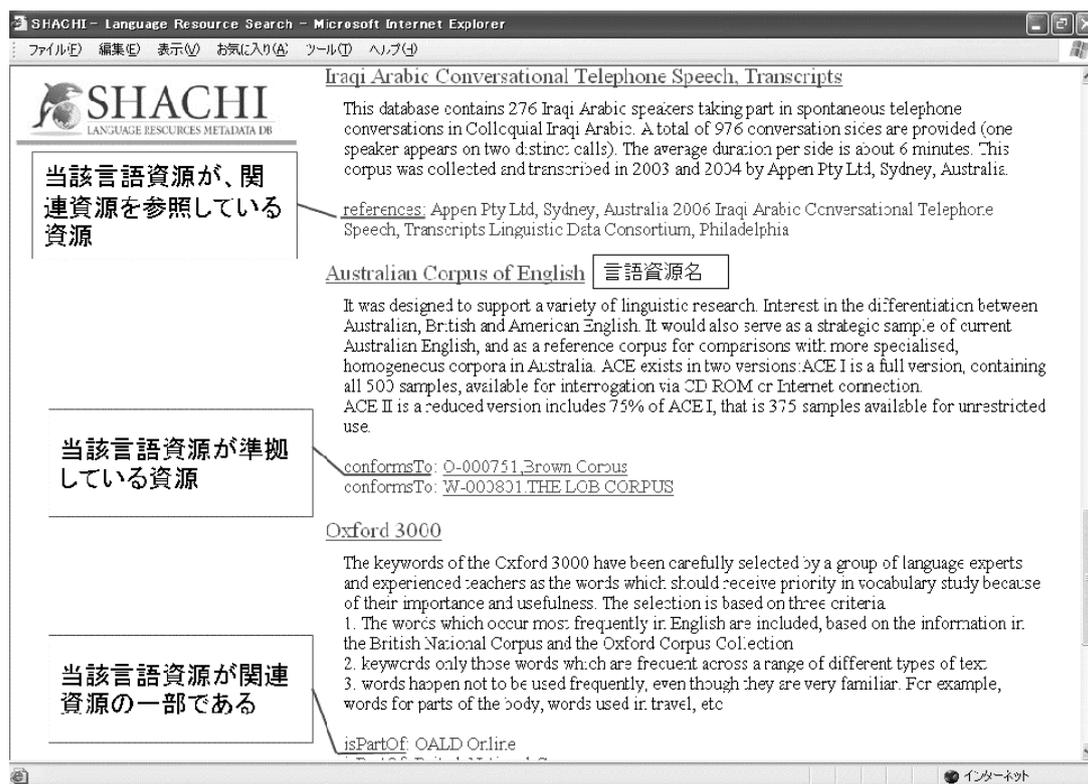


図3. 2. 2-2 言語資源間の関係性の明示化 (SHACHI 検索結果の表示)

## SHACHIの設計

言語資源のメタデータを体系的に蓄積しアクセス性を高める試みとしては、上述したOLAC以外にも、ISLE Meta Data Initiative (IMDI)、DFKIが運営するLanguage Technology World などがある。一般に、情報技術の進展ならびに社会への還元を促進する上で、複数のコンソーシアムが相互に連携して活動することが重要である。SHACHI のメタデータは、OLAC のメタデータセットに準拠しており、それを拡張する形で、より詳細なメタデータを収集している。これは同時に、ダブリン・コア (Dublin Core) のメタデータに準拠していることを意味しており、メタデータの蓄積・流通に適した設計となっている。また、現在、言語属性を示すメタ項目は、国際標準化機構 (ISO) のISO639-2[3] に準拠している。また、日付と時刻の表記に関してはISO8601に準拠している[4]。

## 言語資源メタデータの収集

SHACHI で収集する言語資源は、下記の全ての条件を満たすものと規定している。

- a. デジタル化された言語資源である。
- b. コーパス、辞書、シソーラス、語彙リストのいずれかである。
- c. 英語で記載された公式Web ページ、もしくは、言語コンソーシアム管理下のカタログを有し、かつ、データが公開されている。
- d. 研究機関、研究者、企業によって作成された言語資源である。

これらの条件を踏まえた上で、認知度が高いもの、言語処理技術の発展の上で、主要であると考えられるもの、大規模であるものを優先的に登録している。また、言語資源の流通促進、言語資源の戦略的開発を行うためには、世界の言語資源のメタデータを網羅的に一箇所に集めておくことが有効である。そこで、SHACHIでは、国内の主要言語資源コンソーシアムを始め、欧米、中国の言語資源コンソーシアムの持つ言語資源メタデータをカバーしている[5]。

## メタデータ拡張

ユーザの曖昧な検索に対し、必要十分な情報を提供するためには、言語資源の属性を示す詳細なメタデータに加え、言語資源の属性の近さや上位・下位関係を体系的に表示したオントロジーを構築することが必要であると考えられる。本データベースは、ダブリン・コアの15の基本エレメントに基づくOLACメタデータセットに準拠しており、さらに言語資源の特徴の記述に必要と判断した、新たなメタデータ19項目を追加している。表2の編みかけ部分はSHACHIが採用しているメタ項目であり、左から、ダブリン・コア、OLAC、右端はSHACHI 独自の拡張項目を示している。また、個々の言語資源が、研究者らによって、どのような局面で、どう活用されたかという用途情報は、ユーザにとって貴重である。そこで、本研究では、構文解析技術を用い、言語資源の用途に関する情報を、学術論文から自動抽出する手法を考案し、提供を試みている [6]。

表3. 2. 2-1 SHACHIのメタデータセット表示)

		Qualifier Qualifiers used for more precise description of the resources	
LEVEL 1	DC Qualifiers	LEVEL 2	
DCMES Element	DC Element Refinements	OLAC Extensions	SHACHI Extensions
1 Title	Alternative		
2 Creator	-		
3 Subject	-	Linguistic Subject (29) [olac:linguistic-field] anthropological_linguistics applied_linguistics cognitive_science computational_linguistics discourse_analysis forensic_linguistics general_linguistics historical_linguistics history_of_linguistics language_acquisition language_documentation lexicography linguistics_and_literature linguistic_theories mathematical_linguistics morphology neurolinguistics philosophy_of_language phonetics phonology pragmatics psycholinguistics semantics sociolinguistics syntax text_and_corpus_linguistics translating_and_interpreting typology writing_systems OLAC-Language extension [olac:language]	mono_multi_lingual (2) monolingual multilingual ResourceSubject (4) corpus dictionary thesaurus glossary
4 Description	Table Of Contents Abstract		Language (of description) Price
5 Publisher	-		
6 Contributor	-	Role [olac:role] (24) annotator *author compiler consultant data_inputter depositor developer editor illustrator interpreter interviewer participant performer photographer recorder researcher research_participant responder signer singer *speaker sponsor transcriber translator	Attribute of *Speaker/Author mother-tongue intonation level age gender
7 Date	Created Valid Available Issued Modified Date Accepted Date Copyrighted Date Submitted		
8 Type	(DC Type Vocabulary)	Discourse Type (10) [olac:discourse-type] drama formulaic_discourse interactive_discourse language_play oratory narrative procedural_discourse report singing unintelligible_speech Linguistic Data Type (3) [olac:linguistic-type] lexicon primary_text language_description	Purpose(4) lexicography analysis developing_technologies education Style (2) speech written Form (2) fixed unfixed Sentence(3) short long mixed Annotation (3) annotated plain Annotation_sample Sample
9 Format	Extent Medium		Encoding Markup Functionality
10 Identifier	-		
11 Source	Bibliographic Citation		
12 Language	-	OLAC-Language extension [olac:language]	
13 Relation	Is Version Of Has Version Is Replaced By Replaces Is Required By Requires Is Part Of Has Part Is Referenced By References Is Format Of Has Format Conforms To		Utilization
14 Coverage	Spatial Temporal		
15 Rights	Access Rights License		

### SHACHIのカタログ検索

本メタデータデータベースのユーザが、目的に合った言語資源カタログに到達できるよう、キーワード検索、及び、ファセット検索の2つの機能を整備している。また、現段階では、人手によって構

築した言語資源オントロジーを用い、オントロジー検索機能を試験的に搭載している。図3に、検索ツールの画面を示す。ファセット検索機能は、SHACHIメタデータセットから15種類の主要メタ項目が選択項目として設置されており、ユーザは、自分の希望する言語資源に近い項目を順に選択し、絞り込んでいくことで、該当する言語資源にたどり着くことができる。一方、オントロジー検索では、言語資源オントロジーをたどる方法により、ユーザのニーズに合致する言語資源に到達するシステムであり、漠然とした目的を持ったユーザや、初心者、ならびに、一般のユーザにとって有効であると考えられる。

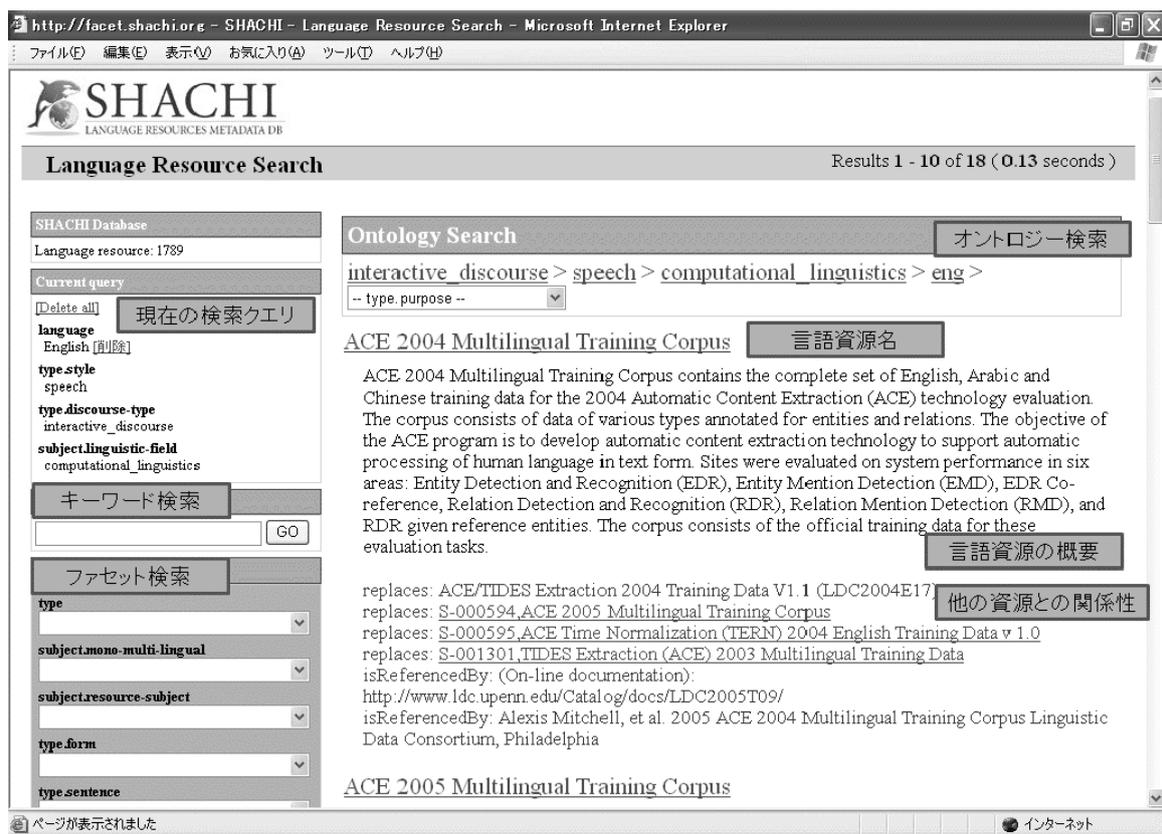


図3. 2. 2-3 SHACHIの検索画面 (http://www.facet.shachi.org/)

## まとめ

情報通信機構（NICT）と名古屋大学では、共同で、言語資源の有効利用、戦略的開発、及び、有機的結合に関する研究を目的に、言語資源のメタデータを大規模に収集している。本稿では、言語資源メタデータベース“SHACHI”の設計、メタデータの収集・拡張、及び、検索機能の実現について述べた。SHACHIはWeb検索によって、認知度の高い言語資源を網羅的に収録しており、また、世界の主要言語資源コンソーシアムが提供している言語資源メタデータをカバーするとともに、さらに、より詳細なメタデータの登録を人手により行っている。現在、約2000件の言語資源メタデータの

登録を完了し、世界最大規模の言語資源メタデータアーカイブとなっている。SHACHIの特徴の1つに、極めて詳細なメタデータを収集している点あげられる。現在、それらの情報を用いて、言語資源のタイプや各言語資源間の近さ（属性の近さ）を計測し、世界中の言語資源メタデータの体系的な蓄積（言語資源オントロジーの構築）を試みている。

#### 参考文献

- [1] T. Ishida, A. Nadamoto, Y. Murakami, R. Inaba, T. Shigenobu, S. Matsubara, H. Hattori, Y. Kubota, T. Nakaguchi, and E. Tsunokawa, A Non-Profit Operation Model for the Language Grid, ICGL, pp.114-121 (2008).
- [2] Y. Hayashi, T. Declerck, P. Buitelaar, M. Monachini, Ontologies for a Global Language Infrastructure, ICGL, pp.105-112 (2008).
- [3] H. Tohyama, S. Kozawa, K. Uchimoto, S. Matsubara, H. Isahara, SHACHI: A Large Scale Metadata Database of Language Resources, ICGL, pp.205-212 (2008).
- [4] ISO639-2: Codes for the representation of names of languages -- Part 2: Alpha-3 code.
- [5] ISO8601:  
[http://www.iso.org/iso/iso\\_catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=4087](http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=4087)
- [6] 小澤俊介・遠山仁美・内元清貴・松原茂樹, 言語資源の効率的利用のための用途情報抽出, 言語処理学会第14回年次大会発表論文集, (2008).

(執筆者 内元清貴)

#### 3.2.3 ヒアリング調査報告I：KOTONOHA計画における著作権処理

第一回目のヒアリングでは、コーパスの構築と著作権処理に関わる問題について、国立国語研究所の前川喜久雄氏に講演いただいた。前川氏は、これまでにも大規模自発音声コーパス『日本語話し言葉コーパス』の開発に中心的に携わった経緯があり、また2006年度からは文科省科学研究費特定領域研究「日本語コーパス」の領域代表者を務めている。「日本語コーパス」では、現代日本語の書き言葉に関する大規模な著作権処理が発生しており、著作権処理の基本方針から、実務に関する詳細についてもヒアリングすることができた。

講演中では、国立国語研究所が推進している現代日本語コーパス整備計画「KOTONOHA」の概要と、その一環として現在構築を進めている『現代日本語書き言葉均衡コーパス』の設計を解説していただき、その中で実施している著作権処理の現状についてご紹介いただいた。本項では、この講演の概要を紹介する。

実施日：2007年10月4日（木）

実施場所：電子情報技術産業協会304会議室

タイトル：「KOTONOHA計画における著作権処理」

講演者：独立行政法人国立国語研究所 前川喜久雄

## KOTONOHA計画

独立行政法人国立国語研究所では、現代日本語の総合的なコーパスを整備する計画として「KOTONOHA」計画を推進している。これは、明治期以降のさまざまな書き言葉・話し言葉を対象として、日本語の総合的なコーパスを順次整備していこうとする事業である。

国立国語研究所では、すでに以下の2つのコーパスを構築・公開している。

- 『日本語話し言葉コーパス』

約661時間、752万語の自発音声（主に独話）を収録し、多くの研究用情報を付与した大規模音声コーパス。2004年公開。科学技術振興調整費開放的融合研究「話し言葉の言語的・パラ言語的構造の解明に基づく『話し言葉工学』の構築」プロジェクト（総括責任者：東京工業大学教授、古井貞熙：研究期間1999-2003年度）による成果。

- 『太陽コーパス』

博文館から刊行された月刊誌『太陽』（1895～1928年）をXML文書化し、研究用情報を付した書き言葉コーパス。総文字数は約1,450万字、記事数は約3,400本、著者数は約1,000人。現代日本語の確立期であり、言文一致完成期の書き言葉を分析することにより、表記法の変遷、文語語法から口語語法への移行の過程、新語の定着の度合いなどを通時的な観点から分析することができる。

KOTONOHA計画は、これら2つのコーパスを含むスーパーコーパスとして設計されている。計画全体は、次の図のとおりである。（[http://www2.kokken.go.jp/kotonoha/ex\\_1.html](http://www2.kokken.go.jp/kotonoha/ex_1.html) より転載）

# KOTONOHA

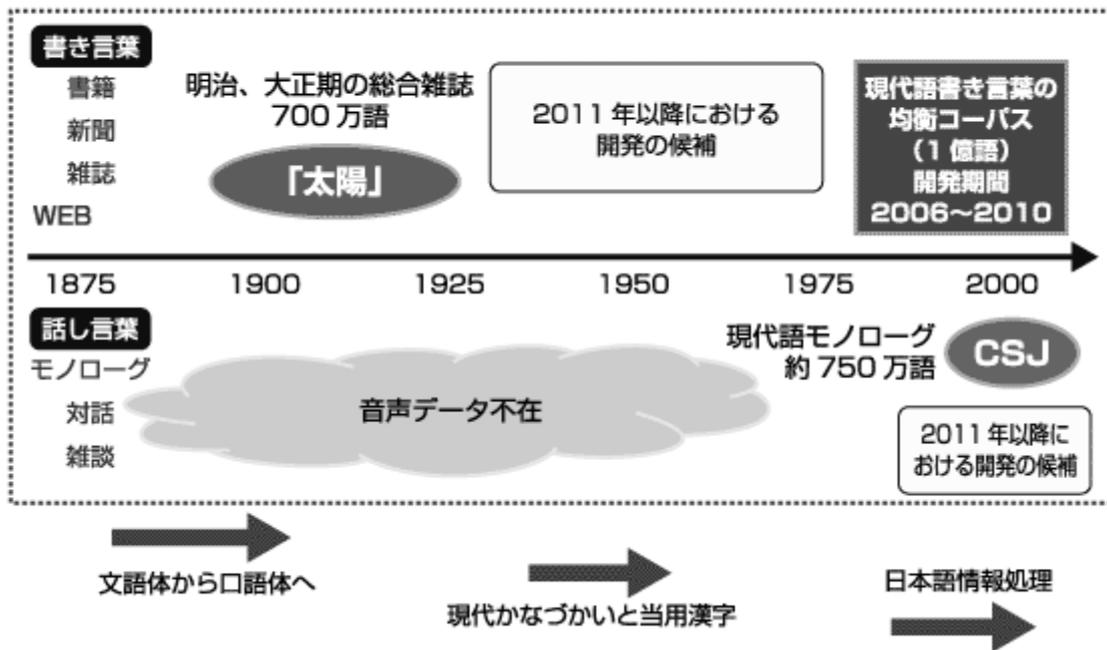


図3. 2. 3-1 KOTONOHA計画

## 『現代日本語書き言葉均衡コーパス』

2006年に、KOTONOHA計画の一環として、現代日本語の書き言葉を対象とするコーパス『現代日本語書き言葉均衡コーパス (Balanced Corpus of Contemporary Written Japanese ; 以下BCCWJ)』の構築を開始した。2006年7月には、この計画の一部が文部科学省科学研究費特定領域研究「日本語コーパス」として採択された。これにより、各大学・研究機関の研究者からなる研究チームを編成し、大規模な体制でコーパス構築を展開している。

BCCWJは、3つのサブコーパスから構成される。各サブコーパスには、書籍・雑誌・新聞・白書・教科書・国会会議録・Web上の文書などから取得した文章 (サンプル) が含まれ、全体で1億語強のサイズが予定されている。BCCWJの全体構成は、以下の図のとおりである。

([http://www2.kokken.go.jp/kotonoha/ex\\_2.html](http://www2.kokken.go.jp/kotonoha/ex_2.html)より転載)



図3. 2. 3-2 BCCWJの全体構成

BCCWJは、その特徴の一つに「著作権を心配せずに使えるコーパス」であることを掲げている。このためには、コーパスに採録するテキストに関して著作権処理を実施し、著作権者からの利用許諾を得る必要がある。BCCWJにおける著作権処理の基本方針は、以下のとおりである。

- すべてのサンプルに著作権処理を施して誰でも使えるコーパスにする
- 権利者には原則として無償利用を依頼する

以下ではこの問題について詳しく述べる。

### BCCWJにおける著作権の在り方

BCCWJに含まれる各サンプルの著作権の在り方には、次の6通りが考えられる。

- 1) 著作権が存在しないもの
- 2) 著作権が存在するが自由に利用できるもの
- 3) 法人が著作権者となっているもの
- 4) 権利は個人に属するが、法人が利用許諾をだせるもの
- 5) 個人が著作権者だが、著作権管理を団体に委託しているもの
- 6) 個人が著作権を保有し、管理しているもの

1) は、法律が最たる例である。著作権法第13条には、著作権法の権利の目的とならない著作物が挙げられている (憲法その他の法令、告示、訓令、通達など)。2) は、「著作権フリーマーク」が付いた白書、国会会議録などが該当する。国会会議録は、国会議員の総意として、自由に利用できることになっている。3) は、白書の大部分、新聞記事の大部分、雑誌の一部、書籍のごく一部などの法

人著作が該当する。4) は、インターネット掲示板やブログ、学术论文の一部などが該当する。

5) は、日本文藝家協会の会員の一部、社団法人日本音楽著作権協会（JASRAC）などが該当する。

6) は、書籍の大部分、雑誌や新聞に掲載された署名記事（連載小説、依頼原稿、投稿など）などが該当する。

このうち、実際に著作権者への接触と交渉が必要となるサンプルの数量として、3) のタイプが3000サンプル程度、4) のタイプが5000サンプル程度、6) のタイプが27,000サンプル程度と見積もっている。

ただし、1つのサンプルの著作権者が複数人に渡るケースもある。例えば、共著の場合、翻訳・翻案の場合、複数の著作権継承者がいる場合などである。さらに、そもそも著作権者の総数が分からないサンプルも少なくない。このような場合、その分、著作権処理の負担が増すことになる。

## 著作権処理の実務

著作権処理の交渉先は、1) 法人、2) 個人、の2つに分かれる。

### 1) 法人との交渉

法人との交渉先としては、新聞社、作家団体、インターネットプロバイダ等がある。このうち新聞社とは、全国紙・ブロック紙・地方紙・通信社などの各社と交渉を開始しており、交渉が完了した新聞社とは順次「覚書」を締結している。また、作家団体としては、日本文藝家協会、日本推理作家協会、日本ペンクラブ等と交渉し、各団体の会員4,200名に対して、著作物の利用許諾依頼状を送付している。現在約2,800名から回答を得ており、許諾率は98%となっている。

### 2) 個人との交渉

個人と交渉する場合には、個人と直接交渉する場合と、間接的に交渉する場合がある。前者は、紳士録・人事興信録・ウェブ・著作権台帳などによって著作権者の連絡先を特定し、利用許諾状を送付する。後者は著作権者の連絡先が分からないなど、直接接できない場合で、出版社などに著者の連絡先の開示を依頼する。

このうち、2) に関して、個人の連絡先が判明して直接交渉に持ち込めるのは全体の4割程度であり、多くの場合は出版社からの協力が必要となる。しかしながら、2005年に個人情報保護法が実施されたことにより、出版社側が著作権者の連絡先を開示することに慎重になっている傾向がある。

## 著作権は処理できるのか

著作権処理を進めていくに従い、最大限の努力をしても著作権者に接触できない（もしくは著作権者を特定できない）ケースが、当初の想定よりも多いことが判明してきた。これには、個人情報保護法の影響が大きいと考えられる。この場合の対処としては、法的には、サンプルの利用をあきらめる

か、文化庁の長官裁定に持ち込むか、という選択肢となる。文化庁長官裁定については、著作権法第67条に、以下のようにある。

（著作権者不明等の場合における著作物の利用）第67条 公表された著作物又は相当期間にわたり公衆に提供され、若しくは提示されている事実が明らかである著作物は、著作権者の不明その他の理由により相当な努力を払ってもその著作権者と連絡することができないときは、文化庁長官の裁定を受け、かつ、通常の使用料の額に相当するものとして文化庁長官が定める額の補償金を著作権者のために供託して、その裁定に係る利用方法により利用することができる。

ただし実社会の中では、便法も行われている。例えば、著作権者が見つからない場合、申し出があった場合にはそれ以上の利用をしない旨を明記した断り書きを付した上で著作物を利用するケースなどである。また、ウェブ上に著作権者の情報を広く求める公告を掲載するなど、従来の新聞広告に代わる情報収集の手段が用いられている。

## 今後の課題

コーパスを構築または利用する側にできることとしては、次の3点が考えられる。

- 1) 著作権データベースの整備を文化庁および著作権管理団体に要求する
- 2) 他人の権利を尊重する
- 3) コーパス（言語資源全般）が本当に社会の役に立つことを疑いの余地なく示す

1) については、著作権者の権利を守るだけでなく、利用者の便宜を図ることも国の仕事であり、そのような要求を利用者の側から発信していくことが必要であると考えます。2) については、コーパスが公開されるまでは作り手側が最大限の努力を払って著作権処理にのぞむすること、公開後に著作権者から接触があったらその意向を尊重し、すぐに対応できる体制を整えておくことが必要である。3) については、かつてのワードプロセッサがそうであったように、いわゆるキラーアプリケーションの開発が有効である。これには、自然言語処理の技術に期待したい。

ヒアリング資料：

前川、知識情報処理技術専門委員会 言語資源分科会 ヒアリング資料（2007年10月）

（執筆者 丸山岳彦）

### 3.2.4 ヒアリング調査報告II：日外アソシエーツにおける言語資源配布の実態

言語資源の配布実態を調査するために、日外アソシエーツの星俊雄氏を招いて以下のようなヒアリ

ング調査を実施した。

実施日：2007年11月6日（火）

実施場所：電子情報技術産業協会303会議室

タイトル：言語資源=国内新聞記事データ頒布の過去・現在・未来

－著作権者と研究者と頒布担当者の本音－

講演者：日外アソシエーツ 企画営業部 星俊雄

日外アソシエーツは主に新聞記事データなどの言語資源の頒布実績のある会社であり、また新聞記事が研究目的として使われるようになった初期の頃から頒布業務を開始している。講師の星氏は、同社でも言語資源の頒布業務の中心的な役割を担ってきた人物である。ヒアリングでは、過去の言語資源頒布の実態や、今後どのような言語資源が求められ、どのような言語資源が利用できる可能性があるかについて御講演いただいた。講演の概要を以下にまとめる。

- これまでの言語資源の頒布実績について

毎日新聞、読売新聞、朝日新聞の新聞記事の頒布を行ってきた。それぞれの頒布開始時期は以下の通り。

- 毎日新聞データ集 1995年9月1日より
- 読売新聞データ集 2004年7月1日より
- 朝日新聞データ集 2004年11月1日より

言語処理の研究分野において、言語資源として最初に使われるようになった新聞記事データは毎日新聞である。日外アソシエーツ社は新聞記事データが配布されるようになる最初の段階からその業務に携わってきている。

- 頒布業務について

頒布に関わる主な業務は以下の通り。

- 言語資源の周知

研究者に対してダイレクトメールを送っている。新しい年度の新聞記事がリリースされる度にDMを送付している。DM以外には雑誌媒体に広告を出している。現在は雑誌「自然言語処理」、雑誌「日本語学」、言語処理学会年次大会の予稿集に広告を掲載している。言語資源の周知には電子メールを使わない。SPAMと間違われるのは好ましくないというのがその理由である。

- 言語資源利用者からの問い合わせに対する回答
- 言語資源利用者に対する確認

書類の書き方に不備があったり、内容に不明な点があったときに確認をする。

- 言語資源の納品（CD-ROM, DVD-ROMの送付）
- 許諾料の請求回収業務

- 販売特性について

新聞記事データ販売の特性として、利用者への個別対応が必要なこと、利用者と著作権者との利用許諾契約の締結を確実にすること、頒布価格が高額であること、などを挙げていた。

- 課題について

- 現在はほとんどの配布先が理工系の研究者であるが、人文系研究者にも対象を広げる必要があると考えている。特に日本語学の研究者を中心に対象を広げていきたい。
- 研究者が退官したときに、その研究者が利用許諾を受けていた新聞記事データを、退官する教官のいた研究室で継続して使ってよいかといった問題がある。現在は、同じ所属の人なら委譲できるという原則に従って委譲手続きを行っているが、ルールを厳密に規定する必要は残されている。同様に、研究者が研究機関を移る際に、新聞記事データを継続して使ってよいかといった問題もある。
- 利用目的の曖昧さへの対応が必要である。現在は、研究目的でしか新聞記事データを配布していないが、研究目的か商用目的かの判断に迷うような曖昧なケースも多い。研究利用目的とは何かを明確に規定する必要が残されている。

- 将来

現在は新聞記事データをパッケージメディア（CD-ROMまたはDVD-ROM）で配布しているが、ネットワークメディアで配布することを検討している。また、言語資源保持者と交渉を行い、配布可能な言語資源の種類を増やすことを視野に入れている。

講演を拝聴したところ、日外アソシエーツ社では新聞記事データの頒布業務にかなり積極的に取り組んでいる印象を受けた。特に、言語資源保持者と利用者との利用許諾契約を結ぶ業務は、個別の対応が必要なことや細かい点を確認する必要があることなどから、実際にはかなり煩雑な作業が必要となる。GSKのような公共的な機関であれ、日外アソシエーツのような営利企業であれ、言語資源の配布・流通を円滑に促進するには言語資源保持者と利用者の仲介を行う組織が必要不可欠であると感じた。

## スポーツ報知新聞記事データの販売の開始

このヒアリングの中で、講演者の星氏と言語資源分科会の委員との間で、新しい言語資源をどのように発掘していくかが話題になった。言語資源分科会では、本ヒアリングの以前から新しいタイプの

言語資源をどのように発掘するかについての検討をすすめており、その一例としてスポーツ新聞の記事が挙げられていた。現在利用可能な新聞記事は、毎日新聞、読売新聞、朝日新聞、日本経済新聞といったいわゆる主要な新聞社の記事である。文体が比較的固く、話題も似通っている場合も多い。一方スポーツ新聞の場合は文体も多少くだけているし、主要な新聞社が取り上げる話題とは違ったスポーツや芸能などの話題が含まれている。一方、スポーツ新聞の場合は言語資源保持者が明確にわかっているため、データ公開の交渉もやりやすいと考えられる。星氏に上記のような言語資源分科会委員の考えをお話ししたところ、ご賛同いただき、新聞記事データの販売を報知新聞社の担当者と交渉していただくことになった。このような経過を経て、スポーツ報知新聞社の記事を研究目的で販売されることとなった。データの概要は以下の通りである。

### スポーツ報知新聞記事データ

掲載年：1998年版から2007年版

記事数：約5万記事／年

販売先：日外アソシエーツ株式会社、日本データベース開発株式会社

発売時期：2008年1月より

言語資源分科会では報知新聞社と具体的な交渉は行っていないが、本ヒアリングがきっかけとなってスポーツ報知新聞記事の販売が開始されたことから、新しいタイプの言語資源の発掘に多少なりとも貢献できたのではないかと考えている。

参考までに2007年度に新たに利用可能になった言語資源を簡単に紹介する。

- スポーツ報知新聞記事データ
  - JEITAマルチモーダル対話コーパス
  - 大阪外国語大学 多言語平行旅行会話文集（サンプル）
- ただし、現在は配布は終了している。
- グーグルWeb日本語Nグラム第1版

スポーツ報知新聞記事データ以外の言語資源はGSKより入手可能である。本分科会では、今後は新しいタイプの言語資源を発掘し、研究者に利用できるような活動に積極的に取り組んでいきたいと考えている。

(執筆者 白井清昭)

### 3.3 今後の展望

本節では、平成20年度以降の言語資源分科会での活動テーマの可能性について述べる。

言語資源のカタログにおいては、アジア言語の情報の収集を進め、カタログの公開を進める。言語資源協会では、検索企業からの日本語Nグラムデータを公開しているが、このNグラムの使われ方、さらには検索ログの研究利用の可能性について検討する。検索ログにおいては、ほかの人がどのような質問を入力したか、どのように検索過程が進んだか、検索結果のうちでどれを見たか、といった情報が必要となろう。このような現在入手困難なデータの活用は、本分科会が考えるべき項目のひとつである。カタログにおいて、どのような言語資源が実在するかは調査されているが、逆にユーザの視点から、どのような資源が求められているかも調査検討する必要がある。これは企業と大学とで異なるかもしれない。求められている言語資源が存在しない場合、小規模なものであれば、JEITAでの開発も視野に入れて活動したい。

ウェブ上のデータが大量になるにつれ、利用時における著作権、肖像権なども検討すべき課題である。ここでは国内の法律問題や現状の調査も必要である。著作権に関する調査においては、日本で言語資源の蓄積あるいは検索のサーバーを立ち上げるときの問題点、成果物の直接利用や、他のソフトへの組み込みなど、特に商業利用における条件の調査が必要である。

言語資源のフォーマットにおいては、自然言語処理の分野における標準化を目指すことも有効であろう。ユーザに使ってもらいやすい言語資源として、検索ツールなどの開発と公開も大切なことである。大規模データに情報を付与し、その解析済みのデータを公開することも、言語資源利用の裾野を広げるためには必要である。この場合、どのようなテキストに、どのような情報を付与したものが必要とされているのかをきちんと調べておく必要がある。また、実際に利用するためには、そのテキストの検索を容易にするツールなどの提供も必要である。

検索に関する研究開発においては、たとえば、検索インデックスの最適化や新語・流行語の調査には、入力されたクエリーそのものが必要であろう。ページランキングアルゴリズムの改良には、どのページをクリックしたかの情報が必要であろう。パーソナル検索の研究にはユーザ情報が、検索支援の研究には一人の連続入力の情報が必要となろう。これらの情報を含むデータを言語資源として利用可能に出来ればと考える。また、キーワードで検索した場合に、どの結果が上に来れば良いのかは、検索の評価において避けて通れない問題であり、その研究のためのデータも有効である。

言語資源利用の実態調査のために、網羅性が完全ではないにせよ、学会での発表で使われているデータの調査も有効と思われる。どのような言語資源があり、どのように使われるかを把握し、将来のユーザに対して、必要な情報を提示することが必要である。

これらの活動を行うと共に、その結果を広く公開することが必要である。ワークショップの開催や、業界としての指針の提供も目指していきたい。

(執筆者 井佐原均)

## 4. 情報アクセス技術

### 4.1 はじめに

すべてのユーザに同一のインタフェースやサービスを提供するのではなく、各ユーザの性質に合わせたきめ細かいサービスを提供する技術に関心が集まっている。このためにはユーザの利用履歴などからユーザ・モデルを構築する技術が重要となる。また、セキュリティの観点から各ユーザのシステムの利用履歴を保存し、それを監視することにより警告を発したり、問題が発生したときには分析して対応策を講じるなどの利用法が考えられる。

このような背景をふまえ、今年度は、システムやサービスなどの履歴情報の収集とその利用技術に関する動向調査をおこなった。

4.2節では、履歴情報を利用するアプリケーションと履歴情報を収集するためのセンシング技術という2つの観点からヒアリングとショールームの見学による調査をおこなった。調査対象は以下の7件である。

#### 4.2.1 「行動ターゲティング広告」(BTA)

- (1) i-media drive (鶴田淳氏、吉田潤氏)
- (2) cyberwing (広屋修一氏)

#### 4.2.2 「推薦システム」(産総研 神瀧敏弘氏)

#### 4.2.3 「情動分析」(NTTラーニング 山内信治氏、夏井誠也氏)

#### 4.2.4 「センサネットワーク：ビジネス顕微鏡」(日立製作所 鈴木敬氏)

#### 4.2.5 「ホームネットワーク」

- (1) 「基礎技術と標準化」(北陸先端大 丹康雄氏)
- (2) 「モデルケース」ショールーム見学(パナソニックセンター、富士通 netCommunity)

4.2.1と4.2.2は履歴情報を利用する事例であり、4.2.3と4.2.4は履歴情報を収集する技術に関するものである。4.2.5はホームネットワークの領域における実装技術・標準化技術とそれを実装して利用した具体例であり、やや近未来的な話題である。これに対して4.2.1や4.2.2の利用例はすでにビジネスに結びついている例である。特に2007年度の国内広告費では、ついにインターネット広告が雑誌広告を上回るという報告<sup>1</sup>が出ており、中でもユーザの行動履歴からユーザの興味を推定し、その興味に合致した広告をそのユーザに対して集中的にうつという「行動ターゲティング広告」も注目を集めている。今後、さらに新しいタイプのインターネット広告が現れる可能性もある。

---

<sup>1</sup> <http://www.dentsu.co.jp/news/release/2008/pdf/2008008-0220.pdf>

4.3節ではインターネットを検索することにより、履歴情報を利用したサービスを網羅性に収集し、それをいくつかの観点から分類した結果について報告する。収集した事例を分析し、分類の軸として、Web上の履歴情報か実世界の履歴情報という対立軸、履歴情報を分析した2次利用かそのままの利用かという対立軸、何に対する行動の履歴であるかという対象の軸などを利用した。

4.4節では、この分野の研究動向を調査するために、ACM Conference on Electronic Commerceで発表された論文から、推薦やパーソナライズに関連する論文7件を選択し、その概要を紹介している。

最後に4.5節では、情報検索関連の著名な国際会議SIGIR2007に委員のひとりが参加したので、その概要と今年度の調査項目に関連の深い、いくつかの発表について報告している。

[徳永健伸]

## 4.2 行動履歴情報の利用に関する技術動向

### 4.2.1 行動ターゲティング広告

#### (1) i-media drive

行動ターゲティング広告（BTA:Behavioral Targeting Advertisement）のリーディングカンパニーの一つである株式会社アイメディアドライブ、鶴田淳氏、吉田潤氏へのヒアリングに基づき、BTAの基本、ビジネスモデル、および質疑応答を紹介する。

#### ① 行動ターゲティングサービスの分類

行動ターゲティング広告は、検索サイトにおいて検索キーワードに関する情報（広告など）を提示するSearch Solution、サイトにおいてPopUpを出す等を行うAdware/Apyware、広告主向けのソリューションで広告主にネット上の広告媒体を紹介するAdvertiser Solution、広告媒体者向けのソリューションで広告媒体に訪れたユーザに最適な情報を提供するPublisher Solution、サイトのテキストや文脈に応じて最適な情報を提供するContextual Solutionとがある。

#### ② 米国行動ターゲティング市場予測

2003年には285百万ドルであったが、2007年には15億ドル、2008年には21億ドルの予算をBTAに利用するといわれている。（iMedia/Ponemon Study 05）2006年のニールセンの調査によると、主要媒体における接触時間、つまりユーザ（特に年収の高い層）のアテンションは、インターネットにシフトしている。

#### ③ 行動ターゲティングとは

ユーザの行動履歴を情報収集用サイトを通して取得し、ユーザをセグメント化する。その後、セグメントに基づき、ユーザがどのページに行くのに関わらずユーザのセグメントに沿った広告を掲出する。つまり、広告を掲出するサイトのコンテンツの内容ではなく、人を基点とした広告手法である。

アイメディアでは、セグメントの作成方法として、情報収集用サイトがあらかじめ区分しておいたサイトを閲覧したユーザをセグメント化するカテゴリーベース、有効と思われるキーワード（就職、転職、仕事など）が含まれたサイトを閲覧したユーザをセグメント化するコンテンツベース、検索窓において特定のキーワードを検索したユーザをセグメント化する検索ワードベースで行っている。

サイトの収益モデルは、ユーザをグルーピングするサイト、広告を掲出するサイトの双方に利益を分配するモデルである。

#### ④ 行動ターゲティング広告のメリット

例えば、車の購入検討から購入までを考えると、トータルに過ぎる時間は約8週間から12週間といわれている。その期間で実際にユーザーが自動車関連サイトに接触する時間は、総インターネット接触時間のわずか3%である。行動ターゲティング広告では、残りの97%の時間にも有効にユーザーにアプローチ可能である。

アイメディアドライブ社での実験によると、1-10回の広告接触と、11回以上の広告接触を比較すると、広告に対する「純粋想起」「助成想起」「オンライン広告認知」で高い効果が見られた。

#### ⑤ 今後について

アドネットワークをさらに広げていく。ポータル、CGM系、新聞系、ISP系、動画系、専門系のサイト等をアドネットワークで結び、広告の在庫管理と配信コントロールをやっていく。

#### ⑥ その他

米国のBTAリーディングカンパニーであるレベニューサイエンス社では、一日に10億以上の行動履歴を取得している。これは、月間で360億imp相当の広告在庫に相当する。

#### ⑦ Q&A

Q：技術的に求めるものは何か

A：たとえば、セグメントの自動生成技術。今はセグメントは手動で作成している。セグメントの粒度をどのようにするかも課題である。

Q：BTA広告を配するのに向いている広告コンテンツはなにか？

A：BTAでは、コンテンツベース広告に比べ、高額商品や日常的に購入する消耗品の広告に向いている。

Q：ユーザーのセグメントは幾つぐらいあるのか？

A：現段階では50ほど。一人のユーザーが複数のセグメントに属する。

#2007末の情報に更新

Q：サイトの滞在時間は利用しているのか？

A：滞在時間は利用していない。閲覧したURLと回数を利用している。

Q：ユーザをセグメント化しすぎることでの弊害はあるか？

A：セグメントを細かくすればするほど、レアなセグメントに属して配信されない広告在庫が増えてしまう。

Q：アイメディアドライブ社の広告ネットワーク（impActネットワーク）にはどれぐらいの加盟サイトがあるか？

A：TV&ラジオ系、スポーツ系、車系、トラベル系、レディース系、ビジネス系などのサイトで、計約170サイト、総UUは約6,000万UU/月、総impは約7億impである。

#2007末の情報に更新

[森田 哲之]

## (2) サイバーウィング

株式会社サイバーウィングの広屋修一氏へのヒアリングに基づき、行動ターゲティング広告（Behavioral Targeting Advertising. 以下、BTAと記述）の市場動向と活用事例を紹介する。

### ① BTAとは

BTAとは、Webサイト上での視聴者の行動履歴を基に、視聴者をセグメント（カテゴリ）に分類し、その視聴者に適した広告を配信するものである。

行動履歴とは、クッキー情報を基にして、一定期間内のサイト訪問、検索、広告反応、購入などの行動を追跡したものである。この履歴を分析することにより、視聴者に適した広告を配信することができる。

### BTAの例) A自動車ディーラーの広告

- ・米国テキサス州のある地方紙サイトにて、車セクションを過去30日間に2回以上訪問した人に対して、別セクションのページでバナー広告を表示した
- ・クリック率が7.7%にも上った（通常の20倍以上）

従来のターゲティング広告は、広告掲載面の内容を分類し、掲載面に適した広告を配信する。それに対し、BTAでは、広告が表示されるサイトの内容と、広告の内容に関わりがなくとも良い。

BTAの広告の対象は、自動車、不動産、求人など、単価が高く、検討期間が長いものが多い。

② 市場動向

米国のBTA市場規模は、2007年には年間15億ドル、2008年には年間20億ドルを超えると予想されている [eMarketer : AdTech,2006.4]。米国の主なBTA企業を表4.2.1-1に示す。

表4.2.1-1 米国のBTA企業

企業	備考
Revenue Science	BTA専業
TACODA	BTA専業
24/7 Real Media	ネット広告事業
Double Click	ネット広告事業 (Googleが買収)
Atlas Solutions	ネット広告事業 (Microsoftが買収)
Advertising.com	ネット広告事業

日本における主なBTA企業を表4.2.1-2に示す。

表4.2.1-2 日本のBTA企業

企業	備考
DAC	Revenue Scienceと提携。impActの行動ターゲティングサービスを発表 (2006/2)
スパイスボックス	2006/6から[X+1] (旧Pontdexter) の販売開始を発表 (2006/5)
Yahoo! Japan	行動ターゲティング、2006/7より販売開始 (2006/5)
サイバーウィング	行動ターゲティング広告配信基盤CW-BTAを販売開始 (2007/3)
サイバーエージェント	MicroAd (旧BlogClick) で行動ターゲティング販売開始 (2006/5発表)

### ③ 日本におけるBTAの事例

以下に日本におけるBTAの事例を挙げる。視聴者はセグメントに分類され、セグメントごとの広告が配信される。

#### ・ Yahoo! 行動ターゲティング

- トップ面以外のスーパーバナーとして掲載
- セグメント：B2B、エンターテインメント、スポーツ、テクノロジー、ファイナンス、医薬、自動車、小売、消費財、通信、不動産、旅行交通、その他
- 単価：0.2円/インプレッション※
- 期間：5日～1か月

※インプレッション：1回表示。BTAでは、この単位当たりの単価が主流。

(Cf. Google AdWordsは、1クリックあたりいくらという単価)

#### ・ BIGLOBE BT広告

- BIGLOBEのサイト閲覧履歴、検索履歴を利用し、バナー広告を掲載
- セグメント：クルマ、住まい、マネー、美容・コスメ、仕事、デジタル機器
- 単価 0.4～0.65円

#### ・ mixi×@Cosme Behavior Match

- @Cosme (コスメ情報サイト) を複数回訪れたユーザに対し、mixi (SNSサイト) 上でコスメ・女性系の広告を掲載
- アイメディアドライブ社が実施

#### ・ 日経BP社のBTメール広告

- 日経ビジネスオンライン (会員登録制) の情報を利用し、広告メールを配信
- 静的属性 (プロフィール) ×動的属性 (閲覧履歴) で詳細にセグメント
- 基本料金：¥60/1通、最低料金：¥500,000/3000件まで (以降1通ごとに¥60)
- セグメント (静的属性1×動的属性1)
  - ◇ 静的属性：年齢、性別、業種、部門、役職、従業員数、年収、興味分野、地域
  - ◇ 動的属性：ニュース、NB100、投資、金融、経営 (戦略)、IT&技術、国際、自己啓発、経営 (人)、ライフ

### ④ BTAの種類

BTAの種類は、たとえば表4.2.1-3のように分類できる。

表4. 2. 1-3 BTAの種類

	Type	説明	ターゲット顧客
1	BT Ad (単独サイト)	あるメディアサイト* 上で行動履歴を取り、同サイト内の広告配信に反映させる	潜在顧客
2	BT Ad Package	メディアサイト 共通カテゴリの複数サイトを束ねて一つの広告製品とする。 (技術的には1のBT Adと同じだが、営業的に違う。小さくて単独では売れない広告を、複数サイト分束ねて、ある程度の大きさで売る)	
3	BT Ad Network	複数のメディアサイトで行動履歴を取り、他のサイト上での広告配信に反映させる	
4	Retargeting	企業サイト上で行動履歴を取り、メディアサイトに移動した場合、その企業サイトに再来訪させるための広告をメディアサイト上に配信する。車種のキーワード等、細かいレベルで販売する。	見込み客／顧客
5	BT On-Site Optimization	企業サイト上で行動履歴を取り、企業サイト内に細かい商材の広告を配信する。	

※メディアサイト：広告を取っているポータルサイト（新聞サイト等）

#### ⑤ BTAの効果

BTAの効果进行调查した事例を2つ紹介する。

##### 1) BIGLOBEと博報堂の行動ターゲティング広告実験

- ・ 広告主：自動車メーカー A社
- ・ 広告媒体：BIGLOBE
- ・ 配信規模：約4000万回のバナー広告配信
- ・ 広告配信期間：3週間
- ・ 履歴の収集期間：広告掲載日以前の4週間分
- ・ ユーザセグメント：200万人のユーザを14種類に分類
- ・ 結果
  - クリック率が向上した（最大4.1倍）。車種等、細かく関心領域を限定するほど向上した。
  - アクティブなユーザを誘導できている。BTAに反応したユーザは、クリック先の広告主サイト内で、能動的に情報収集している傾向が見られた。
  - 広告配信後も継続して広告主サイトを訪問している。通常のパナー広告と比べて最大5.8倍。

## 2) 日経BPとNECの行動ターゲティング広告実験

- ・ 広告主：NECビジネスサイトへの入会促進
- ・ 広告媒体：日経BP社ビジネスサイト「Nikkei BP net」
- ・ 配信規模：60万回のローテーション・バナー広告配信
- ・ 広告配信期間：1週間
- ・ 履歴の収集期間：広告掲載日以前の4週間分
- ・ ユーザセグメント：「企業・経営面」と「IT面」の来訪頻度の掛け合わせで細かく分類
- ・ 結果：
  - ・ 一見さん層に比べ、ヘビーユーザ層では最大2.8倍のクリック率を達成した。
    - ◇ ただし、クリック率はフリークエンシー（広告表示頻度）に大きく左右される。
  - ・ フリークエンシーのコントロールをしないと、いい結果は出ない。
  - ・ ヘビーユーザでも、同じ広告を10回出すと、一見さん層と変わらなくなる。

## ⑥ BTAの今後

BTAの今後としては、動画への拡張が検討されている。拡張の一つとして、動画コンテンツの視聴履歴を利用することが考えられている。また、動画コンテンツ内への広告配信も考えられている。

## ⑦ 主な質疑応答

- ・ Q：BTA広告を掲載するメディアサイトとは、AdSenseを入れるようなブログ等も含む？  
A：ありえる。ブログ事業者がブログ内のユーザ行動をとって、ブログ上に広告を出すなど。
- ・ Q：(④の分類BT On-Site Optimizationについて) パーソナライズドページのように、ページのレイアウトをユーザに応じて変更することもある？  
A：レイアウトの変更はない。イチ押しの広告だけを変えていく。
- ・ Q：ターゲットを絞りすぎると、新しい顧客を逃すジレンマがあるのでは？  
A：そのとおりで、バランスの問題。解決策のひとつとして、明らかに違う人はずす、というアプローチもある。ITオタクすぎる人には、ビジネス広告は向かない、など。
- ・ Q：ユーザを分類する際、各セグメントは独立で、依存関係はないのか？  
A：細かくセグメント化しすぎると、販売ボリュームが取れないので、現在は踏み込んでいないが、だんだんセグメントの組み合わせのニーズは出てくると思う。
- ・ Q：BTAなのか、ランダム表示広告か、見る側から判別できる方法はあるか？  
A：基本的には判別できない。しかし、同じ広告に追いかけているようになると気がつくので、追いかける感がないレベルで、と顧客（広告主）に提案している。1人あたりの頻度をコントロー

ルするが、感じ方には個人差がある。

・Q：様々なBTAの会社があったときに、どういうことが強みになる？

A：ルール（自動も人のノウハウも）とデータ。データ分析力。

・Q：Googleだとキーワードで値段が変わる。BTAではセグメントによって値段が変わるのか？

A：不動産は高いなどはある。基本的に、値段は、単価と表示回数で決まる。

[志賀 聡子]

#### 4.2.2 推薦システム

産業技術総合研究所の神寫敏弘先生へのヒアリングに基づき、推薦システム技術の背景、基本的な分類などについて紹介する。なお、各種推薦アルゴリズムの詳細については、人工知能学会誌に3回にわたる連載<sup>1-3)</sup>で紹介されているので参照されたい。

##### (1) 推薦システム登場の背景

大量の情報発信を可能にする社会の高度情報化、情報発信の低コスト化、記憶媒体の低コスト化や通信の高速化による「情報過多 (Information Overload)」により、情報を参照できる状態にあるにもかかわらず、それを利用できない状況を打破する必要性が生じたことが、推薦システム登場の背景として挙げられる。

##### (2) 他の技術との関連

推薦システムと関連する技術として、情報フィルタリング (Information Filtering) やマーケティングの関連技術が挙げられるが、情報フィルタリングは逐次的に入力される情報から、利用者プロフィールに適合するものを選別する技術であり、内容ベースの推薦システムと類似しているが、利用者プロフィールがクエリで表されることが多い、必要なものを見つけるより、不要なものの除外が主な目的であるといった特徴がある。また、マーケティングの関連技術との比較でいえば、マーケティングは供給側の観点での技術、推薦システムは消費者側の技術である、推薦システムでは、マーケティングで必要な顧客層分析などを目的とした全体傾向のレポート作成などは不要であるなどの違いがある。

##### (3) 推薦システムの分類

推薦システムは、一般に言われるパーソナライズに限らない広い範囲を含み、推薦の個人化の度合いによって非個人化、一時的個人化、永続的個人化に分類することが可能である。また、目的によっても概要推薦、利用者評価、通知サービス、関連アイテム推薦、緊密な個人化などに分類することが

可能である。

#### (4) データの入力

推薦システムへのデータの入力としては大きく明示的な獲得と暗黙的な獲得があり、明示的な獲得の場合、正確さ等は増すが、データ量が少ないという課題がある。また、評価値の正確さについては、利用者の評価には同じアイテムを評価した場合にも揺らぎがあることが実験的に確かめられている。

#### (5) 嗜好の予測

嗜好の予測とは、利用者の嗜好データや、アイテムの情報を用いて、利用者の各アイテムへの関心や好みの度合いを予測することである。予測の評価値としては、予測の誤差が基本的な尺度となるが、それ以外の指標として被覆率、セレンディピティ（目新しさ、意外性）などを高水準に保つことが、場合によっては必要となる。

#### (6) 内容ベースフィルタリングと協調フィルタリング

推薦システムでは推薦に使う手がかりによって大きく違う二つの方法がある。一方はアイテムの特徴を利用する内容ベースフィルタリングであり、もう一方は他人の意見を利用する協調フィルタリング。両者ともにドメイン知識の要否や少数派の利用者に対する性能など得意不得意があり、現在はこれらを組み合わせたハイブリッド型が主流となっている。

#### (7) 推薦の提示

予測評価の高いアイテムでも、推薦すべきでないアイテムがある。推薦の提示の際にはこれらをルールベースのフィルタリングによって除外することは必須である。例えば、既に提示したアイテム、上下巻組の書籍などで下巻を選択した場合の上巻、あるいは季節に合わない季節商品などは除外すべきである。セレンディピティの強化のためには正解率を犠牲にしても異なったタイプのアイテムを推薦に加える、などの戦略をとる必要がある。また、推薦理由の提示は、推薦の精度に関わらず満足度向上に重要で、嗜好が似ている人の評価やシステムが持っている確信度を提示するなどの手法がとられる。

#### (8) 主な質疑

Q：推薦システムの計算量についてはどう考えたらよいか、実用上どのように解決しているのか

A：アルゴリズム的には並列化可能な部分が多いためなので大規模システムでは並列処理を導入している。

Q：推薦は消費者側の仕組みということだが必ずしも提供者側の意図が入らないとはいえないのではないか。

A：MovieLensのチームの実験では意図的にランキングを操作した実験では利用者が操作されたことに気がつくので難しい。サクラ攻撃などでユーザの評価が影響される場合はある。

Q：学術的な研究の要素と運用の要素があるが、研究要素として推薦システムに残されたトピックは何か？

A：予測精度については終わっている。セレンディピティなど新しい評価軸について研究がおこなわれている。予測精度は推薦システムの一部しか評価していない指標で、予測精度が同じでもアルゴリズムが違えば満足度は異なる場合がある。購買とか満足につながったかなどは単に予測精度だけからは計れない。セレンディピティの重要性については一致しているが、現段階ではその評価方法が定まっていないため、意外だと思う、ということの数値化する必要がある。現時点では直接に意外性を狙うよりも推薦の幅を広げるなどの方向性で研究が行われている。

Q：現状、実用という観点でいうとどの程度実用化されているのか。

A：アマゾンなど比較的大手では実用化されている。インターネット系以外ではTV番組推薦のTiVoなどが著名。

Q：POSデータのようなものを対象にした推薦システムはどのくらい使われているのか。

A：現状は、推薦というより相関分析などデータマイニング系の技術が分析に使われているのではない。

[神谷 俊之]

#### 4.2.3 情動分析

生体反応の計測結果を活用するサービスの一例として、NTTラーニングシステムズ株式会社の山内信治氏、夏井誠也氏へのヒアリングに基づいて、Emotional分析ソリューションについて紹介する。

(1) 「Emotional分析ソリューション」(情動計測)とは

- ・情動計測技術とは、ディスプレイに表示されたコンテンツ(映像、Web画面など)を被験者(人間)が見たときの目の動き(瞳孔の動き、瞬きの頻度、視点の動き)を計測し、被験者の関心度・注目度の分析を可能にする技術である。
- ・株式会社VIS総研の技術を活用している。

- ・利用範囲は、映像コンテンツ、Webコンテンツ、教育教材、各種デザイン評価など多岐にわたる。
- ・上記の目の動きは無意識下での反応であり、意識的な制御は困難である。目の動きを計測することにより、被験者の反応を客観的に捉えることができる。

## (2) 計測方法

- ・フェイストラッキングカメラにより目の位置を追尾し、アイトラッキングカメラにより目の動きを計測する。計測結果はネットワーク経由でサーバに送信され、サーバ上で解析する。(図4.2.3-1)

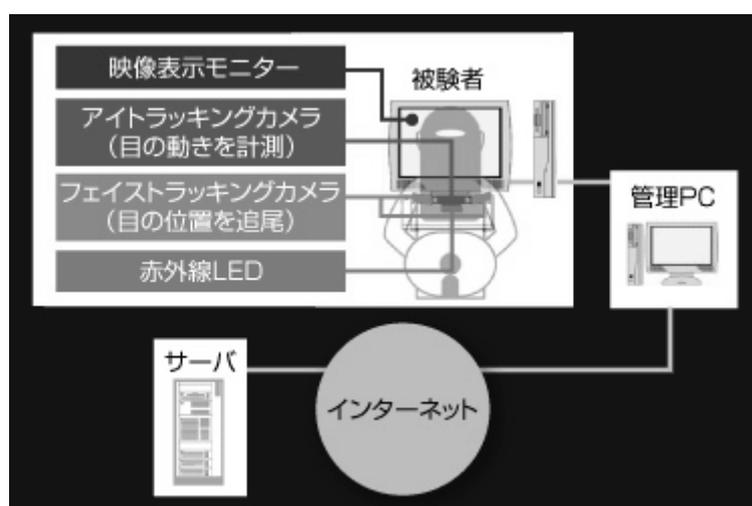


図4. 2. 3-1 Emotional分析ソリューションの計測時の図

- ・計測結果はコンテンツ上にビジュアルに表示される。視線滞留時の瞳孔の径に応じた大きさの円を表示することにより動画コンテンツ上での興味対象とその関心度を分析したり、マウスクリックと視線移動との相関を分析することによりWeb画面のデザインを評価したりすることができる。
- ・医学用途での需要が高い他社ソリューションとの違いは、被験者の自由度の高さと計測の速度である。(他社はアゴのせ型の計測が主)

## (3) 適用事例

- ・Feeling Catcher (コンテンツ評価ソリューション)

映像表示用モニターにテレビ映像などの動画コンテンツを表示し、被験者の視点の動き、瞳孔のサイズ、瞬きの回数を計測する。元の動画コンテンツに視点の動き、注目点(瞳孔サイズ等に基づく興味関心度を円の大きさで表現)を重ねて表示することにより、被験者が対象コンテンツのどの部分に興味を示しているか分析できる。

・ Top Assist (教育・研修)

安全運転講習時の教育教材として使用される「動画映像の危険予知シミュレーション映像」を使用した情動計測により、受講者の危険予知の生データの記録による検証を行い、教材の目的・講習内容の効果を測定し、検証することが可能となる。

・ WebLyzer (WEB評価)

本ソリューションでの情動計測により、各ページにおいてユーザーがどこを多く見ていたか、どこに注目していたかが明らかになり、視点とマウスカーソルの動きによりクリックするまでの行動履歴が明らかになる。

・ Honesty One (アンケート評価ソリューション)

被験者への選択式アンケートと併用する。アンケート回答の信頼度(明確な選択結果かどうか)が取得できる。

・ Trend Catcher (市場動向評価)

車の画像をデザインなどで関心の高い部分を抽出できる。(選択式アンケートと併用)

(4) 主な質疑応答

Q. 情動と感情の違いは何か? たとえば好きなものを見て瞳孔が開くのは感情ではないのか?

A. 好きなものを見て瞳孔が開くのは情動である。嫌悪、好意に関する説明は難しいが、サンプリングデータは増えている。

Q. 頭の動きの許容範囲はどれくらいか?

A. 12cm~15cmくらいである。休憩して席をはずしても再調整なしで続行できる。

Q. 調整時にどのような情報を登録するのか?

A. 被験者の目の形状等を登録する。

Q. 計測速度はどれくらいか?

A. 1/60秒フレームに追従可能な速度である。(計測結果の分析は後処理で別途実行)

Q. 画面輝度との関連性を考慮した正規化はされているのか?

A. 現状はそのような観点での正規化はしていない。通常は同一場所で実施するので部屋の明るさは均一である。音の影響は大きいと考えている。同一コンテンツを複数視聴する場合の正規化も難しく、今後の課題と考えている。

Q. Webページ評価の場合、瞳孔の大きさと視線の滞留時間の有意差はあるのか?

A. 瞳孔の反応大小、視線の滞在時間、視線の軌道などの測定結果で分析、評価をおこなっている。

Q. 写真が好印象を与えたかどうかの判別はできない?

A. 興味関心度の計測はシステムで行い、好意、嫌悪に関してはアンケートで情報収集している。

Q. Webページで瞳孔反応を利用する意味はあるのか？

A. 計測結果をどう活用できるかの検討は次の課題と考えている。

Q. 個人プロフィールとの関連性に関する知見はないか？

A. 瞳孔反応の良し悪しがある。

Q. 処理時間は？

A. 静止画の処理が数十秒、動画はもう少し処理時間が必要。

Q. 瞳孔が大きく、瞬きが少なくなれば興味が大きいという仮説は既存研究で検証されている？

A. 心理学等で研究されている。

[相川勇之]

#### 4.2.4 センサネットワーク

タイトル：日立のビジネス顕微鏡

講師：鈴木 敬様（日立製作所）

##### (1) 研究の目的

- ・『知識労働者の生産性向上』：見えているようで見えていない組織の問題を見える化
- ・『暗黙知領域の展開』：ITだけでなく、直接的な人と人のつながりを促進

##### (2) 研究の経過

- ・センサネットとして始めてから5年が経過
- ・情報端末は10年で100分の1に：初期のPC（1980年代、10<sup>10</sup>）→携帯電話（1990年代、100cc）  
→センサノード（2000年代、1cc）

##### (3) なぜ“顕微鏡”か？

- ・見えなかったものを見る
- ・人間の行動の限界を広げる

##### (4) センサノードの概要

- ・名刺大サイズ：首からぶら下げる名札タイプ。装着時に裏返ると計測できない。
- ・無線ユニット：各所に設置されたアンテナとの相対位置を計測
- ・液晶画面：モノクロ
- ・赤外線センサ：送／受信がペアとなった4つのセンサにより、広角で相手を検知

- ・電池：20.8時間
- ・加速度センサ
- ・音センサ：音量のみ感知

#### (5) 見える化のためのツール

- ・対面グラフ：ユーザどうしが「いつ」、「誰」とコミュニケーションをとっていたのか（実際には対面していたのか）を時系列に表示。加速度センサとの組み合わせにより、どちらからアクションを起こしたのかもわかる
- ・組織タペストリ：加速度の周波数を高・低により色分けして時系列表示。行動のパターン／リズムを可視化
- ・組織地形図：ユーザ間の対面の頻度・時間を元に、コミュニケーションの活発度をネットワーク表示。各ノードがユーザ、アークがコミュニケーションの有無を表す。体面時間が長いほど全体表示の内側に配置される。
- ・五省角：人・行・体・心・知の5つの観点によりビジネス顕微鏡を身に付けての行動を評価。ビジネス顕微鏡の効果についてのフィードバックに利用。

#### (6) 試行実験

- ・日立総合計画研究所のある部署25人にセンサを身に付けてもらい、部長・課長・一般社員のコミュニケーションの違いを計測。
- ・部長・課長は部下と満遍なく、一般社員どうしが密に、といった予想通りの結果が得られた。現在、より詳細に解析中。

#### (7) ライフ顕微鏡

- ・脈・体温・加速度を常時計測
- ・ライフ・タペストリー：観測データを時系列表示。生活リズム・パターンを振り返る。
- ・成功例の目的関数の設定が一番困難
- ・「動き」と「時間」のコラボレーションのマイニング

#### (8) Q&A

Q：チャットやメールなどもっとITツールとの関係を考慮した方がいいのではないか

A：会社内のメール解析には困難が多いのでやっていない。現在MITと共同研究しており、メール等でのコミュニケーション効果については、MITが研究を進めている。

Q：組織地形図をどう活用していくのか

A：いまだ研究段階ではあるが、組織のあり方のようなものが見えるようになり、自己評価のためのツールに発展できればと考えている。

Q：成功事例を元に、生産性が向上する組織コミュニケーションのモデルケースを構築していくべきだと思うが、明確な成功事例は持っているか

A：現状、明確な成功事例は有していない。成功事例の定量的な定義が難しいのが原因の1つにある。ただ、組織の生産性が低下したケースは把握し易く、いくつかの事例を検証済みである。

Q：ビジネス顕微鏡とライフ顕微鏡の組み合わせは？

A：今はしていないが、可能。

Q：位置情報を利用した研究は？

A：そういった研究はあるようだが、詳しくは知らない

Q：音声認識技術の活用は？

A：マイクの性能が良くないので、会話の内容まで認識するようなことをするつもりはない

Q：他にどんなセンサが必要と思うか

A：対面以外（例えば並列して歩きながらの会話など）でのコミュニケーションの有無をセンシングする仕組みがあればと思う

[佐藤 祐介]

#### 4.2.5 ホームネットワーク

##### (1) 基礎技術と標準化

北陸先端科学技術大学院大学および情報通信研究機構（NICT）の丹康雄先生へのヒアリングに基づき、ホームネットワークの基礎技術と標準化について紹介する。

##### ① ホームネットワークとは

ホームネットワークとは、家庭の活動を支援するネットワークシステムである。人間の活動のすべてを含み、様々な年齢層・性別・嗜好の人々をユーザとするので、多種多様な機器・サービスが混在

することになる。そのため、カテゴリ（業界）に分けた分類、整理を行うと同時に、全体をまとめあげるアーキテクチャが重要となる。

ホームネットワークのアイデア自体は昭和60年代から存在し、新しい点は動画の伝送が加わったぐらいで、ほぼ現在と同じような構想が当時から語られてきている。

次世代のホームネットワークシステムは、宅内に配置された数多くのセンサからの情報を利用して、環境や人間の活動のコンテキスト（場所、時間、その他いろいろな状況や嗜好）を認識し、それに適した形でサービスを提供する、いわゆるコンテキストアウェアなシステムとなると考えられている。

## ② 基礎技術

ここでは、次世代ホームネットワークにおいて重要となる基礎技術について、最新の取り組みをいくつか紹介する。

### ・ Floor Markerによる位置推定

地下街のような場所で、通路に2次元バーコードを印刷したものをところどころに配置しておき、ユーザの手元のカメラ付の端末でそれらを撮影することにより現在位置を推定することができる。実データでリアルタイム処理する実験を実施している。推薦システムとの連動も可能である。

### ・ 電波と動画を使った位置推定とトラッキング

指向性の強いアンテナを通路に一定間隔で配置しておき、アンテナの真上にユーザの利用している端末が来れば位置が特定できる。同時に監視カメラからの映像を用いてユーザを同定し、トラッキングすることができる。

### ・ 二次元通信システム

面で構成された通信媒体を用いる。媒体面に端末を近接させるだけでギガビット級の高速通信が可能となる。さらに、媒体から端末への電源供給も同時に行うことができる。二次元通信シートとコネクタからなる。

ホームネットワークでの応用例を考えると、天井に配して各種センサが連動するようしたり、テレビラック上にシートを配してその上に置くだけでテレビが見られたり、書斎のデスク上ではPCと関連機器がワイヤレスで電源不要で連携したりする仕組みが挙げられる。

## ③ 標準化

上記のような基礎技術を含むホームネットワークの要素技術は、互換性と相互運用性の観点から見て標準化される必要がある。

国際標準化に関する標準化団体には以下のようなものがある。

- ・ ITU-T (通信)
- ・ ISO/IEC JCT-1 SC25 (情報機器) … HGI, ECHONET等に関する標準化活動
- ・ IEC TC100 … DLNA, ECHONET等に関する標準化活動

これらの国内対応団体としてTTC, ARIB, JEITAなどがある。

また、システムレベルでのフォーラムとしては以下のようなものがある。

- ・ HGI (Home Gateway Initiative) … 主に通信事業者観点からのHGW,サービス等
- ・ CableLabs … DOCSISなど、CATV関連
- ・ UOPF … 主に宅間接続関係
- ・ DLNA … 主にAV機器関係
- ・ ECHONET … 主に白物家電、家庭向け住宅設備

また、これらの調整をするような役割の団体としてNGN-Forum ホームネットワークWGがある。

さらに、要素技術レベルでのフォーラムとして以下のようなものがある。

- ・ DSL Forum … HGW関連、リモートメンテ等
- ・ OSGi … ソフトウェアアーキテクチャ
- ・ SPIA … ミドルウェア関連

ほかに、各種データリンク、物理媒体として、PLC、Ethernet、WLAN、センサ、POF、可視光通信、等を取り扱う団体も標準化には関連がある。

#### ④ 質疑応答から

##### Q. ホームネットワークとプライバシーについて

A. 作る側よりも案外ユーザ側は受け入れ可能なのかもしれない。アメリカで広く利用されるようになったTiVoがそうだった。Gmailにしても同様。

その点に関して家電系は保守的と言えるだろう。IPv6にHOMENETの人たちが乗ってこなかったのは、つながりすぎる技術が嫌われたこともある。ゲートウェイをつけて（意図的に）つなげないとつながらない方が好ましいと思われるようだ。

[池野篤司]

## (2) 「モデルケース」ショールーム見学

### ① 概要

#### (a) 日時

平成19年8月28日 (火) 14:00-17:30

#### (b) 場所

・パナソニックセンター EUハウス（有明）

・富士通 netCommunity（幸町）

(c) 出席者（敬称略）

徳永（委員長、東工大）、佐藤（日立）、相川（三菱電機）、中臣（リコー）、志賀（富士通研）、  
續木（松下）

② パナソニックセンター EUハウス

(a) 概略

(7) EUとは、Ecology & UDの略。EcologyとしてHEMS（Home Energy Management System）：創エネ、省エネを、UDとして安心・安全を提案する。

(i) 2010年のくらしを想定し、「家まるごと」を提案する。

・住宅棟の想定家族 …… 夫婦＋娘＋祖母（4名）

・フロア数 …… 地上4階（ペントハウス付き）

・建築面積（住宅棟＋運営等） …… 315.87m<sup>2</sup>（96坪）

・主体構造・工法 …… 鉄骨造（耐震構造・重量鉄骨NSラーメン構造）

・軒高 …… 7.343m

・延床面積（住宅棟のみ） …… 1・2階合計 262.00m<sup>2</sup>（79坪）

(ii) フロアは、9つのエリア（①玄関、②エネルギーコーナー、③バス・サニタリー、④キッチン、⑤リビング、⑥ホームシアター、⑦寝室、⑧ホームオフィス、⑨ウェルネスコーナー）から構成されている。

(b) 玄関

・虹彩認証システム



図4.2.5-1 虹彩認証システム

- ・防犯システム：壁を乗り越えようとする壁の上のセンサが反応し、警告を出す。壁の上のものが人間なのか動物（猫など）なのかは、センサに対して与える振動によって判別している。



図4. 2. 5-2 防犯システム

- ・風かもめ：風力と太陽光を使ったハイブリッド型発電システムで、底に付いてる照明とネットワークカメラの電源を供給している。最大190W発電することができ、5日間位は、天気が悪くてもOK。

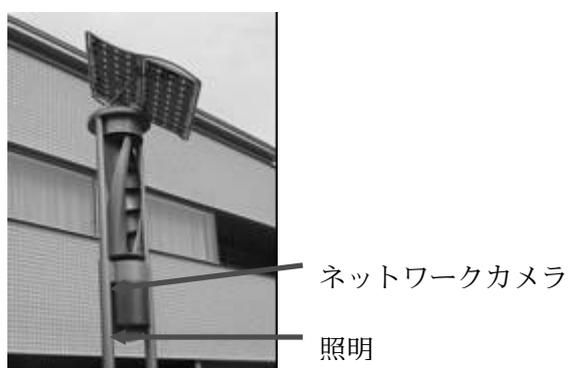


図4. 2. 5-3 風かもめ

- ・スマートエントリーシステム：RFIDタグ付きの携帯電話で家族を識別して、ドアを自動で解錠する。

#### (c) エネルギーコーナー

- ・家庭用燃料電池コージェネレーションシステム：ガスから電気とお湯を作るシステム。EUハウスで電気代6万円、ガス代15万円かかる。ランニングコストとしては、年間5万円の節約になっている。

- ・ HEMS：あまり電気が使われていない時間帯に洗濯機や食器洗い乾燥機を使うなど、端末から効率のよい電気の使い方などが提案される。



図4. 2. 5-4 HEMS

(d) バス・サニタリー

- ・ 酸素ぽっかぽか美泡湯：ユーザが設定したモードで入浴環境を設定。酸素でできたマイクロの泡で、リラックス効果を出す。TVの映像も設定できる。



酸素の泡

図4. 2. 5-5 酸素ぽっかぽか美泡湯

- ・ ゲンキ浴シャワー：朝の場合、生体のリズムに合わせた照明とシャワーで瞬時に体を温め、体を覚醒させる。シャワーで使用する水の量は従来の半分。



図4. 2. 5-6 ゲンキ浴シャワー

(e) キッチン

- ・キッチンナビ:キッチンにおける情報取得や家中の機器のコントロールを行うことができる。  
訪問者の応答、レシピ表示、テレビ・ラジオ視聴などもできる。

(f) リビング

- ・プラズマウォール:大画面プラズマTV (103インチ) を情報の窓「エンターテインメントの窓」、「コミュニケーションの窓」、「セキュリティ&エコロジーの窓」、「アンビエンスの窓」として、使用する。



図4.2.5-7 プラズマウォール

- ・エンターテインメントの窓: VODサービスのデモ
- ・コミュニケーションの窓: 各自の部屋にいる家族と通信
- ・セキュリティ&エコロジーの窓: 玄関のセキュリティセンサの以上検知や訪問者履歴の閲覧、  
登下校中の子供位置をネットワークカメラなどで確認



図4.2.5-8 プラズマウォール（セキュリティ&エコロジーの窓）

- ・アンビエンスの窓：机の上に携帯電話を置くと、そのユーザに合わせてPDPのI/Fをカスタマイズする（例：おばあちゃん→文字フォント大きく）

(g) ホームシアター

PDP、Blu-rayディスクレコーダ、7.1chサラウンドシステム、照明、カーテンなどを一元管理して操作するデモ

(h) 寝室

照明、ベッドの傾き、音楽、PDPに映す映像などを自動制御して快眠、または、快適な起床へ誘う。

(i) ホームオフィス

個人認証システムとWeb会議システムでホームオフィスを実現する。



図4. 2. 5-9 ホームオフィス

(j) ウェルネスコーナー

- ・ジョーバクラブ：エクササイズマシン「ジョーバ」と連動して、やる気を起こさせるコンテンツがPDPに表示される。



図4. 2. 5-10 ジョーバクラブ

- ・音声認識リモコン：「おはよう」、「おやすみ」と話しかけると、照明やカーテンの開け閉めが自動で行われる。



図4. 2. 5-11 音声認識リモコン

### ③ 富士通 netCommunity

#### (a) 狙い

- ・新しい社会のしくみと新しいIT利活用を提案する

#### (b) RFIDの活用

スーパーマーケットを事例に説明

- ・ユーザ側のメリット：携帯電話での買い物进行を想定。食品のトレーサビリティ情報やレシピ情報が提示される。
- ・業者側のメリット：陳列数の把握が容易。陳列数が少なくなると供給するように警告される。



図4. 2. 5-12 スーパーマーケットにおけるRFIDの活用紹介

#### 実験システムの紹介

- ・秋葉原の検品作業：時間が10分の1になった。
- ・小学生の登下校管理：ランドセルにRFIDを仕込み、校門で入出門チェックする。立教小学校（導入から3年）、浦和ルーテル学院（導入から2年）他で導入済み。
- ・IC図書館：蔵書にRFIDを貼付する。蔵書の棚卸などの管理業務や貸出・返却処理を効率化し、職員の作業軽減と利用者サービスの向上を実現。
- ・豚肉トレーサビリティ：豚の耳にRFIDタグを付ける。生産者が、良質の肉が出来た要因分析が容易になる。消費者としては、豚肉のトレーサビリティが可能になる。
- ・フューチャーストアプロジェクト（経済産業省）：三越銀座店と共同で実証実験を行う。ユーザが持っているアクティブタグでユーザの来店、位置情報がわかる。商品にタグを仕込むこ

とで、陳列している商品に希望の商品が存在するのか？バックヤードに存在するのか？を店員が直ぐに把握することができ、接客の質が向上する（存在しない商品を薦めない）。また、試着回数などのデータを分析することで、接客や商品の課題を抽出することができる。実験期間は2週間程度。今年度は、化粧品売り場で実験。

(c) ヘルスケアへの取り組み

「自分の健康は自分で守る」をテーマにコンセプトイメージアニメを作成

- ・ベッドにセンサを付けておき、寝返り、いびきなどから眠りの深さを推定
- ・朝、鏡の前に立つと健康診断が行われ、パーソナルエージェントからアドバイスが与えられる
- ・腕時計型のウェアラブルセンサーを装着し、体温、脈拍、ユーザの位置をセンシング。万が一、倒れても、位置情報から直ぐに救急車両が駆けつける。
- ・ユーザの体調に合わせた食事メニューの提案
- ・メンタル情報を24時間蓄積することで、ユーザの体調を管理。体調やユーザのスケジュールに基づいて、健康アドバイスを行う。（例：最近、会議中に血圧が高いので、休憩をとった方がよいですよ）
- ・ダイエット中の場合、目標体重に向けた健康プログラムを作成
- ・1日の消費カロリーと摂取カロリーも管理



図4.2.5-13 コンセプトイメージアニメ

在宅ケアシステムの紹介

- ・自治体を中心に多数導入されている。（兵庫県社町の事例を紹介。）
- ・独居老人の健康をチェック
- ・高齢者は血圧などを毎朝計り、サーバが管理
- ・保健師がデータから高齢者の状態を判断し、健康アドバイスを高齢者側にPC画面を通じて提

示

- ・緊急通報ボタンもあり、高齢者が押した場合、センターに通報。家族にもメールが届く。

(d) 駅の案内板システム

RFIDタグが仕込まれた杖（高齢で足の悪い人が持っていることを想定）をシステムが検知した場合、ガイド情報の文字フォントを大きくしたり優先席の場所を案内したりする。英語で書かれたガイドブック（外国人観光客が持っていることを想定）を検知した場合、ガイド情報が英語表示になる。

(e) 記録媒体持ち出し管理

記録媒体にRFIDタグが貼付されており、持ち出し管理を行う。記録媒体へのアクセス権判定のための個人認証としては、手の平静脈認証を行う。



図4. 2. 5-14 記録媒体持ち出し管理システム

(f) UBWALL (ユビウォール)

公共の場所や店舗などの人が多数集まる場所に置かれる情報提供PDPシステム。大画面で多くの人に興味を持ってもらい、近づいてきた人にはパーソナライズされた情報やサービスを提供する。PDP上でタッチ操作して、ユーザが欲しい情報を探し出す。携帯電話と連携することで、個人に適応した画面I/Fに自動的に変更可能。また、ユーザが欲しい情報をPDP（サーバ）から携帯電話に移すことも可能。



図4.2.5-15 UBWALL (ユビウォール)

[續木 貴史]

## 4.3 履歴情報を活用したサービス事例

### 4.3.1 調査の概要

本節では、履歴情報を活用したサービス事例の調査について報告する。本調査では、Web上のサーチエンジン（Google、Yahoo）を用いて「履歴、サービス」「履歴、提示」という2種類の条件で検索し、それぞれの検索結果の上位100件から目視で履歴活用サービスの事例と思われるものを抽出するという手順で80事例を収集した（調査は2007年6月時点の検索結果をもとに実施）。

収集した事例を履歴活用の対象アイテムと履歴の活用手法によって分類したのが表4.3.1-1である。履歴活用の対象アイテムでは、対象を特定しない推薦エンジン自体や各種の商品に汎用的に使われる推薦エンジン本体と、それ以外の事例内で言及されている個別アイテムの分類を行った。また、履歴活用方法では、インターネット上のサービス事例と、実世界データをも取り込んで利用する事例を分類した。

履歴活用対象アイテムとしては、Web上の様々な情報、Webページから、書籍、音楽/ビデオなどネットショッピングでよく扱われる商品群、さらに化粧品やレストランなど幅広い対象に及んでいる。また、インターネット上のサービス・エンジンでは、まず履歴情報を加工・分析して推薦に用いる場合と履歴情報を単純に分類するなどして利用者に再度提示する場合を分け、さらに履歴表示のデータが個人に閉じている場合と、複数利用者間で共有される場合を分類して、それぞれについて利用する履歴の種類によって細分類を示した。

表4.3.1-1では、履歴情報を推薦に用いるサービスはAmazon、iTunesなどのサービスが主である。推薦ではなく情報を共有する、関連するアイテムを示すという形での非分析・複数利用者の履歴活用サービスが最も多く、Webページ利用（ブックマーク）、書籍、音楽の購入履歴の共有や、レストラン、旅行、健康ダイエットといった領域で口コミ情報の共有サービスが盛んに行われていることが読み取れる。また、実世界のデータを活用した事例はその特性から移動・行動コミュニケーションや機器情報の活用などで実験的な事例が見られるようになってきているが、実用的なサービスとして供されているものはまだ比較的少数であることがわかる。

表4.3.1-1 履歴情報活用サービス事例の分類

		インターネット上のサービス・エンジン						非ウェブ/実世界/ユビキタス			
		分析(推薦)			非分析(履歴表示)			実世界情報活用	トレーサビリティ		
		(複数利用者の履歴)			個人の履歴	複数利用者の履歴					
		ウェブ操作・検索・購読履歴	機器操作履歴	購入履歴	ウェブ	所有・購入・利用履歴	レビュー・投稿			リアルタイム/最新/ホット	
推薦エンジン本体		(A-1)KBMJ (A-2)SBI (A-3)ALBERT (A-4)教えて一な (A-5) Visionalist				(E-1) Socialtunes (E-2) オクトモ	(F-1) ExpoTV (F-2) Wize (F-3) ThisNext (F-4) Yahoo!Tech (F-5) 価格.com		(H-1) NetCommunity (H-2) 東大ライフログ (H-3) 東大知能住宅 (H-4) PASMO (H-5) リアリティー・マイニング	(I-1) 食品トレーサビリティ (I-2) キュービートレーサビリティ	
履歴活用対象アイテム	単語・キーワード/情報					(E-3) コトノハ	(F-6) Digg (F-7) オリラン (F-8) Kaboodle (F-9) Yedda (F-10) Work.com	(G-1) goo (G-2) 瞬!ワード (G-3) kizasi.jp (G-4) Blogmill (G-5) Blog Ranking (G-6) Blog360			
	Webページ/RSS	(A-6) Google (A-7) SocialFeed			(D-1) gooメモリ・リリーバ (D-2) Google/パーソナライズ検索	(E-4) AttenTV (E-5) del.icio.us (E-6) Buzzurl (E-7) MM/memo (E-8) LVEMARK (E-9) Snippy (E-10) はてなブックマーク (E-11) ニフティクリップ (E-12) livedoorクリップ					
	書籍			(C-1) Amazon		(E-13) フクログ (E-14) 本棚.org					
	音楽/ビデオ			(C-2) iTunes		(E-15) 音ログ (E-16) Last.fm (E-17) PLAYLOG	(F-11) YouTube	(G-7) YouTube (G-8) テレビ王国			
	化粧品						(F-12) @COSME				
	レストラン						(F-13) Yelp (F-14) 食べログ (F-15) askU.com (F-16) Yahoo!グルメ (F-17) B-Restaurant (F-18) ぐるナビ				
	旅行・ホテル						(F-19) フォートラベル (F-20) じゃらん (F-21) 楽天トラベル				
	住宅情報	(A-8)ネクスト									
	就職情報						(F-22) 就職活動日記				
	TV番組		(B-1) AdapTV							(H-6) パナソニックユビキタスハウス	
	健康・ダイエット						(F-23) ダイエット日記 (F-24) Yahoo!ビューティー (F-25) オムロンヘルスカウンタ (F-26) ダイエット日記 (F-27) からだカルテ		(H-7) NTT NOTE		
	OA機器								(H-8) DocuHouse (H-9) PC監査サービス		
移動・行動コミュニケーション								(H-10) マイライフアシスト (H-11) ビジネス顕微鏡 (H-12) On Demand Bus (H-13) D混む.jp (H-14) インターナビ・フローティングカーデータ (H-15) スマートループ			

表 4.3.1-1に示した各事例の概要については次節で紹介する。

[神谷 俊之]

### 4.3.2 調査結果

#### (1) ウェブ操作履歴

No.	項目	概要
A-1	KBMJ <a href="http://www.kbmj.com/">http://www.kbmj.com/</a>	Webサイトを訪問したユーザーの行動履歴をもとに推奨商品を表示するASPサービス「パーソナライズド・レコメンダー」を販売。 (関連記事) <a href="http://itpro.nikkeibp.co.jp/article/NEWS/20060324/233225/">http://itpro.nikkeibp.co.jp/article/NEWS/20060324/233225/</a>
A-2	SBI VeriTrans <a href="http://www.veritrans.co.jp/recommend_asp/">http://www.veritrans.co.jp/recommend_asp/</a>	「レコメンド ASP サービス」を提供。基幹システムとの連携をとらずに、タグで行動履歴を取得して実現するため、コスト、時間、手間をかけずに導入できることを特徴とする。
A-3	ALBERT <a href="http://www.albert2005.co.jp/">http://www.albert2005.co.jp/</a>	リコメンデーションエンジン「Bull's eye」を開発。買ったさとスペックの関係を「コンジョイントモデル」の応用最適化した独自の商品データベースと、その個人が理想とする商品と、推薦する商品との距離測定方法に特徴がある。
A-4	「教えてーな」 <a href="http://www.oshiete-na.com/">http://www.oshiete-na.com/</a>	上記(A-3)のALBERT社が運営する消費者支援サイト「教えて！シリーズ」の総称であり、同じ商品を持っている人や、その商品を欲しいと思っている人が情報交換できるSNS（ソーシャルネットワーキングサービス）の一種である。「教えて！シリーズ」には、以下の4種類がある。 ・教えて！家電 ・教えて！Beauty ・教えて！ブライダル ・教えて！保険
A-5	Visionalist <a href="http://www.visionalist.com/">http://www.visionalist.com/</a>	ROI（Return On Investment）向上のためのWeb解析ツールである。サイト内におけるアクセスユーザーの典型的な行動を知るための「行動履歴分析」機能などをもつ。同機能を用いた分析により、サイト運営者が想定している「導線」と、ユーザーの実際の「動線」を知る事ができ、導きたいページへユーザーを誘導するための対策をうつことができる。
A-6	サイト推薦機能  米Google <a href="http://www.google.com/">http://www.google.com/</a>	米Googleが提供する機能である。「検索履歴」機能の利用者に対して、それに基づいてユーザーが好みそうなサイトを推薦する。新しいサイトを見つけるために検索語を入力しなくてもよいという特徴がある。 (関連記事) <a href="http://internet.watch.impress.co.jp/cda/news/2007/04/19/15472.html">http://internet.watch.impress.co.jp/cda/news/2007/04/19/15472.html</a>
A-7	Social Feed  リアルコム社 <a href="http://www.realcom.co.jp/">http://www.realcom.co.jp/</a>	RSSリーダである。ユーザーの趣向や購読履歴を利用し、おすすめの記事を抽出する。 (関連記事) <a href="http://www.realcom.co.jp/news/20070611_SocialFeed.html">http://www.realcom.co.jp/news/20070611_SocialFeed.html</a>
A-6	注文住宅ポータルサイト 「家づくりHOME'S」  株式会社ウィルニック <a href="http://www.willnic.co.jp/">http://www.willnic.co.jp/</a>	住宅・不動産情報ポータルサイトである「家づくりHOME'S」において「レコメンドサービス」機能を提供している。同サイトを訪れたユーザーの「アクセス数」や「お気に入り登録」、「資料請求」、といった行動履歴を統計解析し、「掲載会社」・「完成事例」・「住宅商品」各情報別に、「この会社を見た人はこちらもチェックしています」といったオススメ情報を提

		示す。KBMJのASP サービス「パーソナライズド・レコメンダー」を使用している。 (関連記事) <a href="http://v.japan.cnet.com/news/release/story/0,2000067550,00018620p,00.htm">http://v.japan.cnet.com/news/release/story/0,2000067550,00018620p,00.htm</a>
--	--	--

## (2) TV番組

No.	項目	概要
B-1	視聴環境適応型表現技術の研究 (AdapTVの研究) <a href="http://www.nhk.or.jp/strl/group/systems/systems03.html">http://www.nhk.or.jp/strl/group/systems/systems03.html</a>	NHK放送技術研究所では、「AdapTV」と名付けられた放送サービスの研究が行われている。放送局から送るコンテンツとコンテンツの内容などの説明情報(メタデータ)を活用し、視聴環境や視聴者の嗜好に応じてコンテンツを知的に変換して表現する技術が研究されている。 応用イメージとして、語学番組における学習レベルにあった番組の提示、ニュース番組における視聴時間に適合したダイジェストニュースの作成などがある。

## (3) 購入履歴

No.	項目	概要
C-1	Amazon <a href="http://www.amazon.co.jp/">http://www.amazon.co.jp/</a>	書籍、DVD、CD、おもちゃ、家電などを販売する通販サイトである。商品説明ページには、協調フィルタリングを用いたオススメ情報(この商品を買った人はこんな商品も買っています)が提示される。
C-2	iTunes Store <a href="http://www.apple.com/jp/itunes/store/music.html">http://www.apple.com/jp/itunes/store/music.html</a>	Apple社の音楽・動画ダウンロード販売サービスである。Just For Youには、曲の購入履歴に従っておすすめ曲が表示される。

## (4) ウェブ

No.	項目	概要
D-1	Googleパーソナライズド検索 <a href="http://www.google.co.jp">http://www.google.co.jp</a>	ユーザーが検索した履歴を表示、管理、学習することで、ユーザーの嗜好に合わせ、より求めている検索結果であるほど上位に表示される。
D-2	gooメモリ・リトリーバβ <a href="http://labs.goo.ne.jp/mr/">http://labs.goo.ne.jp/mr/</a>	Webブラウザのアクセス履歴を保存・解析して、後からキーワードなどで検索できる行動履歴の管理を支援するツールである。

## (5) 所有・購入・利用履歴

No.	項目	概要
E-1	Socialtunes <a href="http://socialtunes.net/">http://socialtunes.net/</a>	自分の持ち物情報を管理・公開したり、友だちと情報交換を行うためのサービスである。興味のある商品、持っている商品、欲しい商品を管理・公開できる自分専用のページをもつことができる。またレビューを書いて公開することも可能である。
E-2	オクトモ	楽天オークションのiモード版で提供されるSNSで、オークションに組み込まれており、商品を通じてユーザーのマッチングを自動的に行っている。

	<p>楽天オークション  <a href="http://auction.rakuten.co.jp/">http://auction.rakuten.co.jp/</a></p>	<p>る。趣味・嗜好の合うユーザー同士、双方の認証を得たユーザー同士で、出品をメールで通知したりグループを作成したりすることにより、取引を促している。</p> <p>(関連記事)  <a href="http://www.nttdocomo.co.jp/info/news_release/page/20061113.html">http://www.nttdocomo.co.jp/info/news_release/page/20061113.html</a></p>
E-3	<p>コトノハ  <a href="http://kotonoha.cc/">http://kotonoha.cc/</a></p>	<p>経験や感じたこと、好き嫌いなどの様々なコト（キーワード）に対して○と×をつけていくことで、自分を表現したり、他のユーザーとコミュニケーションを行ったりすることができる。○×をつける対象はコトと呼ばれ、コトノハユーザーであれば自由に登録することができる。単語やキーワードだけでなく文章の登録も多い。コトに対する○×とコメントを通しての緩やかなコミュニケーションが行われているのが特徴。</p>
E-4	<p>AttenTV  <a href="http://www.atten.tv/">http://www.atten.tv/</a></p>	<p>他のユーザーが訪問しているウェブサイトのビジュアルデータを提供するサービスである。</p> <p>(関連記事)  <a href="http://japan.cnet.com/special/media/story/0,2000056936,20348376,00.htm">http://japan.cnet.com/special/media/story/0,2000056936,20348376,00.htm</a></p>
E-5	<p>del.icio.us  <a href="http://del.icio.us/">http://del.icio.us/</a></p>	<p>ソーシャル・ネットワーキングを加味したウェブ・ベースのブックマーク。自分のブックマークのほか、他の人のブックマークを自由に取り込むことができる。</p>
E-6	<p>Buzzurl  <a href="http://buzzurl.jp/">http://buzzurl.jp/</a></p>	<p>ECナビラボが2007年1月10日にリリースしたソーシャルブックマークサービス。</p>
E-7	<p>MM/memo  <a href="http://1470.net/mm/">http://1470.net/mm/</a></p>	<p>shinao氏によるソーシャルブックマークサービスの一つ。2005年1月12日サービス開始。タグに加えて、クリップしたURLへのコメントを記すことができる。</p>
E-8	<p>LiVEMARK  <a href="http://livemark.jp/">http://livemark.jp/</a></p>	<p>ソーシャルブックマークサービスの一つ。2005年2月8日サービス開始。参照元を記録する機能に特徴がある。</p>
E-9	<p>Snippy  <a href="http://snippy.jp/">http://snippy.jp/</a></p>	<p>ウェブデザイナー向けのソーシャルブックマークサービス。招待制。</p>
E-10	<p>はてなブックマーク  <a href="http://b.hatena.ne.jp/">http://b.hatena.ne.jp/</a></p>	<p>はてなのソーシャルブックマークサービス。</p>
E-11	<p>ニフティクリップ  <a href="http://clip.nifty.com/">http://clip.nifty.com/</a></p>	<p>ニフティ版のソーシャルブックマークサービス</p>
E-12	<p>livedoorクリップ  <a href="http://clip.livedoor.com/">http://clip.livedoor.com/</a></p>	<p>livedoorが2006年6月27日にリリースしたソーシャルブックマークサービス。</p>
E-13	<p>ブックログ  <a href="http://booklog.jp/">http://booklog.jp/</a></p>	<p>他人が所有している本を本棚形式で提示</p>
E-14	<p>本棚.org  <a href="http://pitecan.com/Bookshelf/">http://pitecan.com/Bookshelf/</a></p>	<p>他人が所有している本を本棚形式で提示</p>
E-15	<p>音ログ  <a href="http://otolog.jp/">http://otolog.jp/</a></p>	<p>他人が所有しているアルバムや現在他人が聞いている音楽がわかる</p>
E-16	<p>Last.fm  <a href="http://www.lastfm.jp/">http://www.lastfm.jp/</a></p>	<p>ユーザーの音楽的嗜好を反映するSNSであり、それを通じてユーザー同士の音楽的なつながりをつくり出す。</p>

		(関連記事) <a href="http://ja.wikipedia.org/wiki/Last.fm">http://ja.wikipedia.org/wiki/Last.fm</a>
E-17	PLAYLOG <a href="http://playlog.jp/">http://playlog.jp/</a>	ソニーが運営。自分がパソコンで再生した音楽の履歴をアップロードする等の特徴がある。

## (6) レビュー・投稿

No.	項目	概要
F-1	ExpoTV <a href="http://www.expotv.com/">http://www.expotv.com/</a>	商品レビューに動画を利用したソーシャルコマースサイト。テキストベースの従来の商品レビューと比べ、非常に強いインパクトを消費者に与えている。 (関連記事) <a href="http://japan.internet.com/column/wmnews/20070522/6.html">http://japan.internet.com/column/wmnews/20070522/6.html</a>
F-2	Wize <a href="http://wize.com/">http://wize.com/</a>	Amazon.com や Shopping.com、Yahoo.com などの大手 EC サイトを含む約7,000のサイトから、129万を超えるユーザーレビューを検索・収集し、レビューの内容を分析するアグリゲートレビューサービス。 (関連記事) <a href="http://japan.internet.com/wmnews/20070522/6.html">http://japan.internet.com/wmnews/20070522/6.html</a>
F-3	ThisNext <a href="http://www.thisnext.com/">http://www.thisnext.com/</a>	ロサンゼルスに本拠地を置くベンチャー企業の ThisNextが運営するソーシャルコマースサイト。多くの消費者が自分のお気に入りの商品にレビューを書き加え、サイト内で共有している。 (関連記事) <a href="http://japan.internet.com/column/wmnews/20070313/6.html">http://japan.internet.com/column/wmnews/20070313/6.html</a>
F-4	Yahoo! Tech <a href="http://tech.yahoo.com/">http://tech.yahoo.com/</a>	米Yahooが提供しているコンシューマー向けの技術サイトである。「サポート」「機能」「使いやすさ」「バッテリー」など9種の軸からレーティングを行うことが可能であり、細かな情報をユーザーから収集することができる。 (関連記事) <a href="http://japan.internet.com/column/wmnews/20070320/6.html">http://japan.internet.com/column/wmnews/20070320/6.html</a>
F-5	価格.COM <a href="http://kakaku.com/">http://kakaku.com/</a>	パソコン・家電から葬儀費用の比較・見積案まで様々なジャンルの商品の比較サービスを提供する、国内最大規模のインターネット比較検索サイトである。 (関連記事) <a href="http://d.hatena.ne.jp/keyword/%b2%c1%b3%ca%2ecom">http://d.hatena.ne.jp/keyword/%b2%c1%b3%ca%2ecom</a>
F-6	Digg <a href="http://digg.com/">http://digg.com/</a>	テクノロジー系のニュースサイトである。ユーザーの投稿が「Digg.com」のトップページを飾るかどうかは、他のユーザーの「票 (digg)」をいくつ獲得したかによって決まる。 (関連記事) <a href="http://japan.cnet.com/interview/story/0,2000055954,20096134,00.htm">http://japan.cnet.com/interview/story/0,2000055954,20096134,00.htm</a> <a href="http://www.jiten.com/index.php?itemid=5359&amp;catid=4">http://www.jiten.com/index.php?itemid=5359&amp;catid=4</a>
F-7	オリラン <a href="http://oriran.com/">http://oriran.com/</a>	利用者がランキングやアンケートを登録し、他のユーザーが投票するというインターネットコミュニティウェブサイト。 (関連記事)

		<a href="http://ja.wikipedia.org/wiki/%E3%82%AA%E3%83%AA%E3%83%A9%E3%83%B3">http://ja.wikipedia.org/wiki/%E3%82%AA%E3%83%AA%E3%83%A9%E3%83%B3</a>
F-8	Kaboodle <a href="http://www.kaboodle.com/">http://www.kaboodle.com/</a>	ソーシャルショッピングサイトである。ユーザーによる投票機能を加えた独特のウィジェットを提供している。 (関連記事) <a href="http://japan.internet.com/column/wmnews/20070410/6.html">http://japan.internet.com/column/wmnews/20070410/6.html</a>
F-9	Yedda <a href="http://yedda.com/">http://yedda.com/</a>	知識共有サイトである。各ユーザーのアドバイス数や「役に立った」という評判の数を、プロフィール欄に集計して表示する。 (関連記事) <a href="http://japan.internet.com/column/wmnews/20070123/6.html">http://japan.internet.com/column/wmnews/20070123/6.html</a>
F-10	Work.com <a href="http://www.work.com/">http://www.work.com/</a>	ビジネス実務に関するアドバイスを共有する知識共有サイトである。各ユーザーのアドバイス数や「役に立った」という評判の数を、プロフィール欄に集計して表示している。 (関連記事) <a href="http://japan.internet.com/column/wmnews/20070123/6.html">http://japan.internet.com/column/wmnews/20070123/6.html</a>
F-11	YouTube <a href="http://youtube.com/">http://youtube.com/</a>	動画共有サイトである。各動画に対して、閲覧数、コメント数、お気に入り登録数、そしてスターレーティング（五つ星による評価）が表示される。 (関連記事) <a href="http://japan.internet.com/column/wmnews/20070123/6.html">http://japan.internet.com/column/wmnews/20070123/6.html</a>
F-12	@COSME <a href="http://www.cosme.net/html/aac/index.html">http://www.cosme.net/html/aac/index.html</a>	化粧品の口コミサイト。口コミの分析やユーザの意見の近い口コミが記載されている商品を検索できる。
F-13	Yelp <a href="http://www.yelp.com/">http://www.yelp.com/</a>	地元の店舗などへのユーザーによる批評を集めたソーシャルネットワーキングサイトである。レストランなどの娯楽情報に関するレビューを集める。情報発信に積極的なレビュアーに“elite（エリート）”バッジを与え、レビューに対する信頼を高める仕組みを導入している。 (関連記事) <a href="http://japan.cnet.com/special/story/0,2000056049,20354282,00.htm?ref=rss">http://japan.cnet.com/special/story/0,2000056049,20354282,00.htm?ref=rss</a> <a href="http://japan.internet.com/column/wmnews/20070123/6.html">http://japan.internet.com/column/wmnews/20070123/6.html</a>
F-14	食べログ.com <a href="http://r.tabelog.com/">http://r.tabelog.com/</a>	レストランとお取り寄せ商品の口コミサイト。
F-15	askU.com <a href="http://www.asku.com/rgj/">http://www.asku.com/rgj/</a>	レストラン口コミサイト。
F-16	Yahoo!グルメ <a href="http://gourmet.yahoo.co.jp/">http://gourmet.yahoo.co.jp/</a>	レストランの口コミも存在。
F-17	B-Restaurant <a href="http://www.jpbest.com/brest/review/">http://www.jpbest.com/brest/review/</a>	レストラン口コミサイト。
F-18	ぐるナビ系 <a href="http://www.gnavi.co.jp/">http://www.gnavi.co.jp/</a>	レストラン検索サイト。口コミも存在。
F-19	旅行のクチコミサイト フォー	旅行の口コミサイト。

	トラベル <a href="http://4travel.jp/">http://4travel.jp/</a>	
F-20	じゃらん <a href="http://www.jalan.net/">http://www.jalan.net/</a>	旅行の口コミサイト、ホテルの口コミも存在。
F-21	楽天トラベル <a href="http://travel.rakuten.co.jp/">http://travel.rakuten.co.jp/</a>	旅行の口コミサイト、ホテルの口コミも存在。
F-22	みんなの就職活動日記 <a href="http://www.nikki.ne.jp/">http://www.nikki.ne.jp/</a>	就職活動に関する情報交換用掲示板サービスや、内定者MLサービス等を提供している口コミ就職活動サイト。 (関連記事) <a href="http://d.hatena.ne.jp/keyword/%A4%DF%A4%F3%A4%CA%A4%CE%BD%A2%BF%A6%B3%E8%C6%B0%C6%FC%B5%AD">http://d.hatena.ne.jp/keyword/%A4%DF%A4%F3%A4%CA%A4%CE%BD%A2%BF%A6%B3%E8%C6%B0%C6%FC%B5%AD</a>
F-23	ダイエット日記～かわるナビ <a href="http://kawaru.jp/">http://kawaru.jp/</a>	リンク&コミュニケーション社が運営しているサイトである。なかなか継続することが難しい生活改善やダイエットを、「自分でかわる」「みんなでかわる」「専門家とかわる」という3つの切り口で応援しているのが特徴。 (関連記事) <a href="http://japan.internet.com/column/wmnews/20061128/6.html">http://japan.internet.com/column/wmnews/20061128/6.html</a>
F-24	Yahoo!ビューティー- 90日間ダイエットダイアリー <a href="http://diet.beauty.yahoo.co.jp/special/promotion/diary_1/index.html">http://diet.beauty.yahoo.co.jp/special/promotion/diary_1/index.html</a>	体重や日々の食生活、運動量などを毎日記録していく“ダイエットをしたい人のためのダイアリー”である。
F-25	オムロンヘルスカウンタ Walking Style <a href="http://www.healthcare.omron.co.jp/product/hj710it_1.html">http://www.healthcare.omron.co.jp/product/hj710it_1.html</a>	ユーザー専用の有料サイト「Walker's index」で万歩計の歩数をサーバに登録することができ、歩数ランキングなどを楽しめる。
F-26	ダイエット日記.COM <a href="http://dietnikki.com/">http://dietnikki.com/</a>	ダイエット支援ブログコミュニティである。体重・体脂肪などのデータを書き込むとすぐにグラフが更新されるダイエット日記を作ることができる。
F-27	からだカルテ <a href="http://www.karadakarute.jp/">http://www.karadakarute.jp/</a>	体脂肪計やヘルスメーター、体組成計などの健康機器を開発・販売するタニタが運営する健康支援サイトである。 (関連記事) <a href="http://itpro.nikkeibp.co.jp/article/NEWS/20070308/264315/">http://itpro.nikkeibp.co.jp/article/NEWS/20070308/264315/</a>

#### (7) リアルタイム／最新／ネット

No.	項目	概要
G-1	goo <a href="http://search.goo.ne.jp/index.jsp">http://search.goo.ne.jp/index.jsp</a>	誰かが検索中キーワードを提示。
G-2	瞬！ワード <a href="http://www.nifty.com/search/shun/">http://www.nifty.com/search/shun/</a>	ホットキーワードを提示。

G-3	kizasi.jp <a href="http://kizasi.jp/">http://kizasi.jp/</a>	ブログ中のホットキーワードを提示。
G-4	Blogmill <a href="http://www.blogmill.jp/">http://www.blogmill.jp/</a>	ブログ中のホットキーワードを提示。
G-5	Blog Ranking <a href="http://blog.with2.net/trend_words.php">http://blog.with2.net/trend_words.php</a>	ブログ中の話題のキーワードを提示。
G-6	Blog360 <a href="http://blog360.jp/">http://blog360.jp/</a>	ブログ中のホットキーワードを提示。
G-7	YouTube <a href="http://youtube.com/">http://youtube.com/</a>	(F-11) 参照
G-8	テレビ王国ランキング <a href="http://tv.so-net.ne.jp/guest/pcportal.action">http://tv.so-net.ne.jp/guest/pcportal.action</a>	テレビ王国はソネットエンタテインメント株式会社が運営する、テレビ番組情報サイトにおけるサービスのひとつである。現時点から1週間先の放送予定番組を、「iEPG録画予約」実績をもとにして、地域別にランキングしている。

#### (8) 実世界情報活用

No.	項目	概要
H-1	富士通 NetCommunity <a href="http://jp.fujitsu.com/showrooms/netcommunity/introduction/concept-demo/">http://jp.fujitsu.com/showrooms/netcommunity/introduction/concept-demo/</a>	ユビキタスショッピング、ユビキタス観光、ユビキタス空港など、RFIDタグ関係のデモ展示がある。
H-2	各種センサによるライフログの解析手法・行動予測の展開と応用 <a href="http://www-db.ccr.u-tokyo.ac.jp/ccr_usr/data/3515.html">http://www-db.ccr.u-tokyo.ac.jp/ccr_usr/data/3515.html</a>	日常の人の行動を詳細に計測することを目的に、様々な視点からの情報を総合的に解析するため、ノイズセンサー・加速度計・気圧計・紫外線計・角速度計・GPS等のセンサーをパッケージ化した行動計測装置を開発。東京大学の大学院工学系研究科の研究成果である。
H-3	知能住宅（東京大学） <a href="http://nsk-network.co.jp/050726.htm">http://nsk-network.co.jp/050726.htm</a>	ベッドやソファの中に埋め込まれた500以上のセンサーによって人の位置と行動などの生活パターンを把握し、状況に応じて必要情報を表示・発信する仕掛けがなされた住宅である。東京大学の研究チームが開発。
H-4	交通系非接触式ICカードを利用した連携サービス・PASMOでつくるサービスネットワーク <a href="http://www.igvpj.jp/activity/02_tokyu/index.html">http://www.igvpj.jp/activity/02_tokyu/index.html</a>	利用が急速に広まっている交通系ICカード（PASMO）を用いて、ショッピングや食事などの現実世界でのサービスと、情報推薦や検索などのネットの世界でのサービスとを連携させ、利用者にとって価値の高いサービスを提供。
H-5	リアリティー・マイニング <a href="http://reality.media.mit.edu/">http://reality.media.mit.edu/</a>	位置情報や通話を記録できるようにカスタマイズした携帯電話の履歴をマイニングする。 (関連記事) <a href="http://hotwired.goo.ne.jp/news/technology/story/20050802302.html">http://hotwired.goo.ne.jp/news/technology/story/20050802302.html</a>
H-6	パナソニックセンターTOKYO ユビキタスハウス「イーユーハウス」	ユビキタスハウス「イーユーハウス」で、TVの設定が家族別にカスタマイズされるデモ等がある。

	<a href="http://panasonic.co.jp/euhouse/">http://panasonic.co.jp/euhouse/</a>	
H-7	NTT NOTE <a href="http://www.ngn-note.jp/02_01_03.html">http://www.ngn-note.jp/02_01_03.html</a>	ネットワーク系のショールーム。プレゼンス機能を使った災害時のポータル、血圧等の情報をリアルタイムに送るヘルスケア介護（映像のみ）の紹介がある。
H-8	富士ゼロックス DocuHouse <a href="http://www.fujixerox.co.jp/soft/docuhouse/syukei04.html">http://www.fujixerox.co.jp/soft/docuhouse/syukei04.html</a>	コピー機、プリンタ関連の印刷枚数集計サービスである。 ・さまざまなジョブデータを利用者、部門、機種別などの単位で集計 ・DV*分析ツールによる用紙の削減 などが特徴。
H-9	SecureOne/クライアントPC簡易監査サービス  <a href="http://www.tis.co.jp/news/2006/pdf/060220.pdf">http://www.tis.co.jp/news/2006/pdf/060220.pdf</a>	監査用のUSB機器をクライアントPCに接続してレジストリデータを取得。TIS側でデータを解析し、監査結果報告書を企業に提出する。 (関連記事) <a href="http://www.atmarkit.co.jp/news/200602/21/tis.html">http://www.atmarkit.co.jp/news/200602/21/tis.html</a>
H-10	マイ・ライフ・アシストサービス <a href="http://www.igvpj.jp/activity/01_docomo/index.html">http://www.igvpj.jp/activity/01_docomo/index.html</a>	NTTドコモが経済産業省の産業振興策「情報大航海プロジェクト」のもとで開発。携帯電話経由で個人の属性やネットアクセス履歴、移動情報、さらには声の抑揚までを収集・分析し、ユーザーが“潜在的に求める”情報やサービスを提供する。 (関連記事) <a href="http://www.itmedia.co.jp/news/articles/0711/14/news083.html">http://www.itmedia.co.jp/news/articles/0711/14/news083.html</a>
H-11	ビジネス顕微鏡（日立）	センサ技術を用いて、組織内のコミュニケーションや活動の状況を地形図のように図面化する組織活動監視システムである。 (関連記事) <a href="http://enterprise.watch.impress.co.jp/cda/software/2007/06/22/10575.html">http://enterprise.watch.impress.co.jp/cda/software/2007/06/22/10575.html</a>
H-12	On Demand Bus <a href="http://www.nakl.t.u-tokyo.ac.jp/demandbus">http://www.nakl.t.u-tokyo.ac.jp/demandbus</a>	乗客のデマンド（希望）に合わせて走行する次世代の交通機関。柏市で実証実験中。
H-13	D混む.jp <a href="http://www.disneycom.jp">http://www.disneycom.jp</a>	東京ディズニーリゾートの混雑状況を、土日祝祭日や連休、給料日、ディズニーイベントの開催、学生の修学旅行や卒業旅行情報、過去10年間のディズニーランドの天気情報などから科学的に分析して予測。
H-14	インターナビ・フローティングカーデータ on Google Earth (本田技研工業) <a href="http://www.premium-club.jp/lab/lab1.html">http://www.premium-club.jp/lab/lab1.html</a>	米Google社の地図表示ソフトウェア「Google Earth」を利用して、インターナビ・フローティングカーシステムで収集した交通情報を公開。 (関連記事) <a href="http://www.designnewsjapan.com/news/200603/30auto_hondagikenkougyou060329.html">http://www.designnewsjapan.com/news/200603/30auto_hondagikenkougyou060329.html</a>
H-15	スマートループ構想 (Pioneer)	各ユーザーから収集した渋滞や施設、ルートなどの情報を専用のサーバーにアップロードしてもらい、統計処理した情報をデータ配信の形でフィードバックするというもの (関連記事) <a href="http://arena.nikkeibp.co.jp/article/news/20060510/116602/">http://arena.nikkeibp.co.jp/article/news/20060510/116602/</a>

(9) トレーサビリティ

No.	項目	概要
I-1	日本のトレーサビリティの現状と動向 <a href="http://kumamoto.lin.go.jp/syokuniku/03.html">http://kumamoto.lin.go.jp/syokuniku/03.html</a>	熊本県畜産協会によるトレーサビリティに関する動向調査報告である。石井食品（株）の事例（商品パッケージに印刷された品質保証番号と品質保持期限を入力すると「使用した原材料の産地」「アレルギー情報」「遺伝子組み換え情報」が検索可能）や、JA全農・広島県本部の試み（携帯電話のモバイルサイトを利用した検索）などが紹介されている。
I-2	【中級】トレーサビリティの基礎から実践まで 最終回 <a href="http://itpro.nikkeibp.co.jp/article/COLUMN/20060718/243507/">http://itpro.nikkeibp.co.jp/article/COLUMN/20060718/243507/</a>	トレーサビリティに関する連載記事のひとつである。キューピーの事例が紹介されている。

[相川 勇之]

## 4.4 推薦とパーソナライゼーションの研究動向

本節では、eコマースに関する国際会議であるACM EC (ACM Conference on Electronic Commerce) で発表された論文から、推薦、パーソナライズに関連する論文7件の概要を紹介する。

### 4.4.1 Recommender Systems in E-Commerce

J.B.Schafer, J.Konstan, J.Riedl (GroupLens Research Project, University of Minnesota)

ACM Conference on Electronic Commerce 1999 pp.158-166

#### (1) 概要

本論文では、推薦システムがE-commerceサイトの売り上げ増強にどのように貢献するかを述べている。推薦システムを用いているサイトを6つ取り上げて、インタフェース・推薦するための技術とユーザ入力データ・推薦の提示法の観点から推薦システムを分類している。最後に、推薦システムの今後の展開についてのアイデアをいくつか提示している。

推薦の仕組みが実際に大手のサイトで利用され始めたころの論文なので、取り上げられているサイトのいくつかは現在では統合されてなくなっていたり、サイトはあるが推薦システムが搭載されなくなっていたりするものもある。しかし、推薦システムと一言と言っても簡単なものから高度なものまで様々なものを含んでいることを概観できる点や、それらをいくつかの観点で分類することにより今後の発展を予見している点は、一読の価値はあると思われる。

なお、データを新しくして改訂された詳細版<sup>4)</sup> がジャーナルに掲載されているので、興味を持たれた方はそちらを参照されたい。

#### (2) はじめに

推薦システムは、以下の理由で E-commerce 増強につながる。

- Browsers into buyers … 見ているだけの人を購買者に変えることができる
- Cross-sell … ひとつ買った人に別のものを薦める
- Loyalty … 顧客との関係を強化することによって、またサイトに来てくれる

対象サイト・推薦システムには以下のものを取り上げる。

- Amazon.com … Customers who Bought, Eyes, Amazon.com Delivers, Book Matcher, Customer Comments
- CDNOW … Album Adviser, My CDNOW
- eBay … Feedback Profile

- Levis … Style Finder
- Moviefinder.com … Match Maker, We Predict
- Reel.com … Movie Matches, Movie Map

### (3) インタフェースの分類

- Browsing … 実店舗で探すのと異なり、多くの専門家の意見が参照でき、実際に探して回らなくてもいい
- Similar Item … 顧客が興味を示した
- Email
- Text Comments
- Average Rating
- Top-N
- Ordered Search Results

### (4) 推薦するための技術とユーザ入力データ

- 推薦するための技術
  - Non-Personalized Recommendations … 他の人が一般的に何と言っているかに基づく
  - Attribute-Based Recommendations … 「こんなのを探している」と入れる
  - Item-to-Item Correlation … ユーザが興味を持っている少数の商品から判断
  - People-to-People Correlation … 協調フィルタリング（購買履歴の共通性を見る）
- ユーザ入力データ
  - 購買データ、Likert (=rating)、Text、Editor' s choice

### (5) 推薦の提示法（推薦をもらうために必要なユーザのアクション）

- Organic Navigation … 商品ページに表示される
- Request Recommendation List
- Selection Options … 「こういうのを教えてほしい」と選ぶ
- Keyword/Freeform

### (6) E-Commerce サイトと推薦システムの今後の展開

- Negative rating の（購買データからの）入手
  - 「持ってるけど好きじゃない」情報

- 返品情報
  - 他の商品よりも詳細に見た = 「他の商品」が少し negativeにレーティングされたのと同じ
  - ・ 推薦ではなくユーザに「説明」するシステム
    - 「あなたが以前好きだった〇〇に似た商品です」
    - 持っていない商品データではなく、すでに持っている商品データを参照する必要あり
  - ・ さまざまなデータの統合利用 … 購買データ、レーティング、保有データ
- など

[池野 篤司]

#### 4. 4. 2 Analysis of Recommendation Algorithms for E-Commerce

Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl

(GroupLens Research Group / Army HPC Research Center)

ACM Conference on Electronic Commerce 2000 pp.158-167

Eコマースの商品推薦における有効な推薦手法として協調フィルタリング (CF) がある。CFとは、他の利用者の意見をもとにして、ターゲット利用者への推薦を行うシステムである。CFにおける重要な課題としては、(1) スケーラビリティ：数千万オーダーのデータ（ユーザ数×ユーザあたりの情報）を実時間で処理する必要がある。(2) 推薦品質：推薦された商品が気に入らないものであった場合、利用者は推薦を利用しなくなる。(その意味ではfalse positiveを減らすことが重要である。) の2点が挙げられる。本論文では、CFのプロセスを “Representation”：利用者が購入した商品をモデル化するスキーマ、“Neighborhood formation”：ターゲット利用者の近傍利用者の決定方法、“Recommendation generation”：近傍利用者からTop-N推薦商品を決定する方法という3タスクに分けて実験的に各段階の評価を行っている。

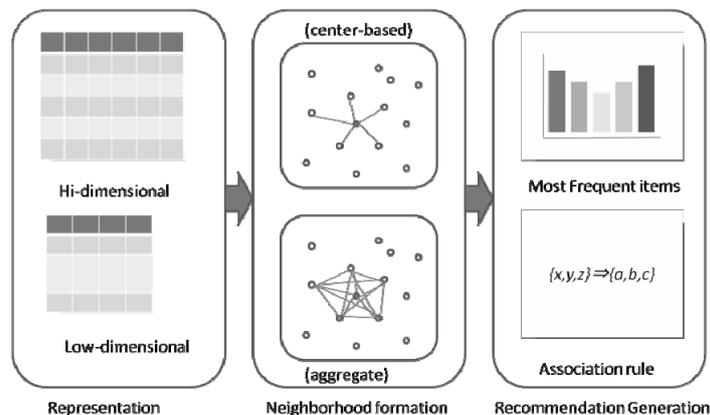


図4. 4. 2-1 推薦システムの3要素

実験用データセットとしては、10万件（913ユーザ×1682タイトルのマトリックス）の映画推薦サイト（MovieLens）のデータと97045件（6502ユーザ×23554アイテム）の実際のEコマースサイトの購入履歴データを用いており、いずれも全体の80%をトレーニングセット、20%をテストセットとして、F1値を用いて評価を行い以下のような点を明らかにしている。

- 1) 推薦を行う際の近傍サイズを大きくすることは一定数まではF1値を改善することに貢献するが、それを過ぎると逆にF1値が悪化する。このためデータセットの特性に応じて最適な近傍サイズを決定する必要がある。本論文ではトレーニングセットをさらに80%、20%に分割してトレーニングセットのみで評価を行うことで、最適値をトレーニングセットのみから決定できることを示している。
- 2) 次元縮退についても同様に次元数を変化させた場合のF1値について検証を行い、同様にテストセットのみから最適な次元数を推定できることを示している。
- 3) 協調フィルタリングによる推薦と従来のルールベースの推薦システムの比較を行い、データ数が小さい場合にはルールベースとの差は小さいが、データセットが大きくなるにつれて協調フィルタリングによってパーソナライズした推薦を行う場合に良好なF1値が得られることなどを示している。

[神谷 俊之]

### 4.4.3 Enabling Scalable Online Personalization on the Web

Debra VanderMeer, Kaushik Dutta & Anindya Datta

ACM Conference on Electronic Commerce 2000, 185-196, 2000.

#### (1) 概要

e-コマースサイトにおいて、過去の訪問履歴と現在の訪問ルートから実時間での動的な製品ナビゲーションを提供するアーキテクチャの開発。ナビゲーション情報を作成する元となる“Profile”を、頻繁に参照されるものほど早くアクセスできるようなキャッシュ構造とすることで既存のデータウェアハウスよりも高いパフォーマンスとスケーラビリティを実現する。

#### (2) システムの構成

ユーザが次に起こしそうなアクションを予測して関連製品へのナビゲーションを提供するe-コマースサイトのシステム（図4.4.3-1）。

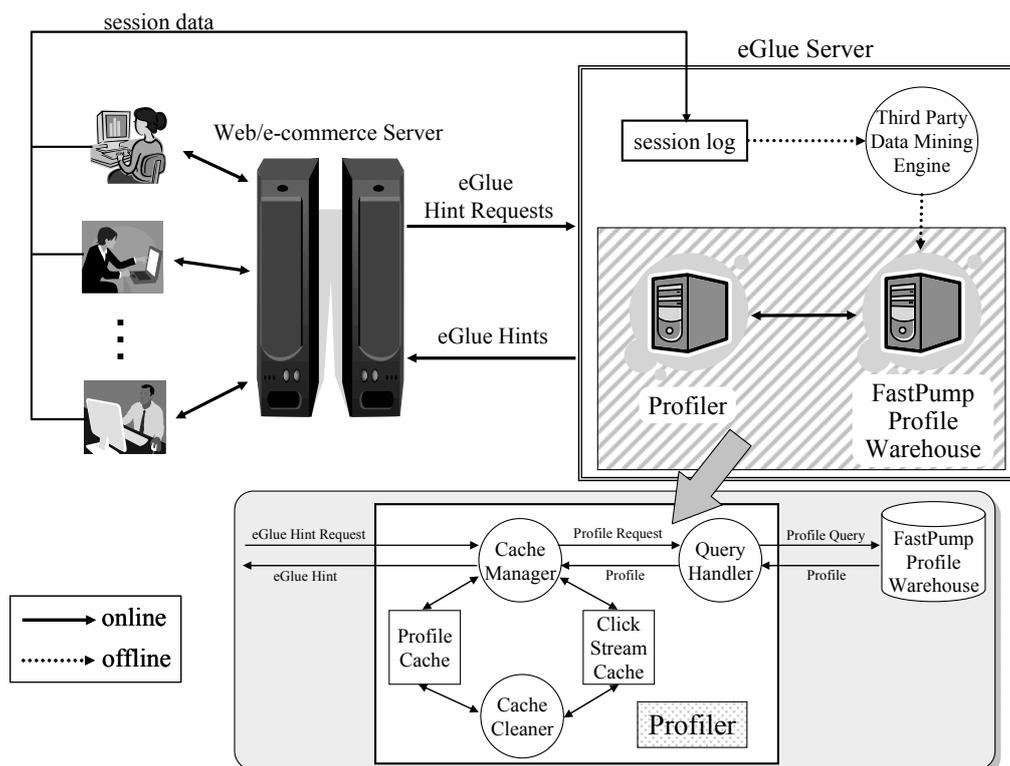


図4.4.3-1 システムアーキテクチャ（斜線部が本論文のコントリビューション）

表4. 4. 3-1 図4. 4. 3-1の各項目の説明

#	名 前 (括弧内は略称)	機 能
1	Web/e-commerce Server(ES)	製品ページとナビゲーション情報を提供
2	session log	ユーザのclickstream履歴を蓄積
3	clickstream	以下の操作を要素とするユーザのアクションシーケンス <ul style="list-style-type: none"> <li>・ navigation : ナビゲーションのためのリンクをクリックする行動</li> <li>・ buy : 購買行動</li> <li>・ departure : サイト離脱行動 (最後のクリックからの経過時間で判断)</li> </ul> 例) <NPapyrus.com, NFiction, NHistorical Fiction, NHF3, BHF3, D>
4	Third Party Data Mining Engine	標準的なシーケンシャルパターンマイニングツール。このツールを用いて session logからさまざまなルールをマイニングしてFPに蓄積。
5	Profiler	eGlue Hint要求を受け取り、clickstreamに基づいてProfile Cache内から該当するprofileを探索。cache内にあればそれを元にhintを生成し、無ければQH経由でFPから新たに生成したprofileを受け取りhintを生成してWESへ送信。
6	FastPump Profile Warehouse(FP)	QHからのprofile要求を受け、clickstreamとアクションルールを用いてprofileを生成
7	eGlue Hint	アクションとその確率のペアから成る
8	Profile	hint生成に用いる図4.4.3-2のフォーマットのルール
9	Cache Manager(CMG)	eGlue Hint要求の監視、PC、CCの更新
10	Cache Cleaner(CCL)	PC、CC内の古いエンTRIESを削除
11	Query Handler(QH)	Profile要求を管理
12	Profile Cache(PC)	Profileをキャッシュ
13	Clickstream Cache(CC)	Clickstreamをキャッシュ

**定義1**

**Profile**は2つの要素  $\langle RA, RC \rangle$  から成る。それぞれ

- ・  $RA$  は長さ  $CSL$  のユーザのclickstream
- ・  $RC$  は帰結  $c_i$  の集合  $\{c_1, c_2, \dots, c_n\}$  であり、各々の  $c_i$  は3つの要素  $\langle A, L, p \rangle$  から成る。それぞれ
  - (a)  $A$  はaction
  - (b)  $L$  は項目ラベル
  - (c)  $p$  はclickstreamが与えられた場合の  $c_i$  の条件付確率

図4. 4. 3-2 Profileの定義

■ 基本動作

- 1) WESが製品ナビゲーション情報をeGlue Hint Request (i番目のクリックとセッションID) としてeGlue Serverへ問い合わせる
- 2) eGlue Server内のProfilerがユーザのclickstreamを元に次に起こしそうな行動を予測した情報をeGlue HintとしてWESに返す
- 3) WESはこのHintを元にナビゲーション情報をWebページ上に表示する

(3) 技術の詳細

FPへのprofile要求の頻度を減らしてシステムを高速化するために、Profile Cache内に有用かつ最新のprofileが保持されるような仕組みを提供。

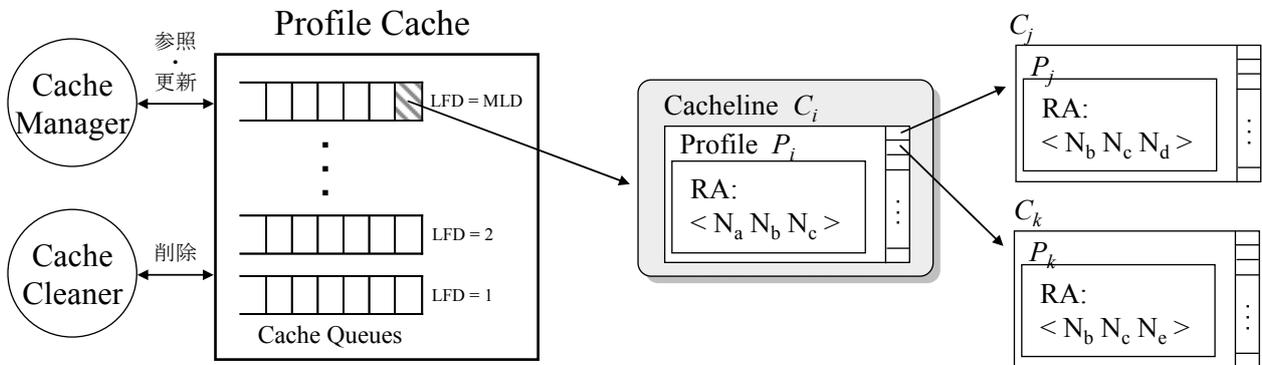


図4. 4. 3-3 Profile Cacheの仕組み

表4. 4. 3-2 図4. 4. 3-3の各項目の説明

#	名前	機能
1	Cacheline(CL)	profileと他のCLへのポインタを要素にもつProfile Cacheのエントリ
2	Cache Queues(CQ)	LFD値ごとに管理しているCLの待ち行列。LFD値が小さいほど優先度が高い。各queue内はタイムスタンプ順に整列されており、最近のものほど優先度が高い。
3	Lowest Following Distance(LFD)	あるCLの集合から対象とするCLへの到達アクション数の最小値 (図4.4.3-4 定義3)
4	Maximum Lookahead Distance(MLD)	先読みするアクションの最大値

**定義2** : *Following Distance*  $F(C_i \rightarrow C_j)$

$C_i.P.RA$ から $C_j.P.RA$ へ到達するのに必要な最小アクション数

例)  $C_i, C_j$ がそれぞれ $P_i.RA = \langle N_a, N_b, N_c \rangle, P_j.RA = \langle N_b, N_c, N_d \rangle$ をもつとすると、 $F(C_i \rightarrow C_j) = 1$ となる。

**定義3** : *Lowest Following Distance*  $LFD(C_i)$

cachelineの集合 $\{C_1, C_2, \dots, C_n\}$ に対して、あるcacheline  $C_i$ の $LFD(C_i)$ とは

$$LFD(C_i) = \min[F(C_1 \rightarrow C_i), F(C_2 \rightarrow C_i), \dots, F(C_n \rightarrow C_i), ]$$

図4. 4. 3-4 Lowest Following Distanceの定義

多くのユーザから参照される可能性が高く、かつ、最近に参照されたものをPC内に保持するようにする。そのための仕組みとして、profile内RAから到達できるRAをもつCLどうしを関連付ける。例えば、あるユーザが $N_a \rightarrow N_b \rightarrow N_c \rightarrow N_d$ 、別のユーザが $N_a \rightarrow N_b \rightarrow N_c \rightarrow N_e$ という順にサイト

内を移動したとすると、図4.4.3-3右のように関連付けされる。

このようなデータ構造の下、ユーザのナビゲーションクリックが発生するたびに以下のような手順でPC内のキャッシュの配置換えを行う。

- 1) ユーザが $\langle N_a, N_b, N_c \rangle \rightarrow N_e$ の順に移動
- 2) CMGは（図4.4.3-3における） $C_k$ を取得しProfile  $P_k$ からナビゲーション情報を提供
- 3)  $C_k$ を現在のqueueの最後尾へ
- 4)  $C_i$ と関連付けているCLのうちMLD以内のもののLFDを計算し、その値に対応するqueueの最後尾にそれぞれを移動

以上のPC内CLの置換処理により、LFD値が高いqueueほど、また、そのqueueの先頭にあるCLほどCCLの削除対象となる。CCLは定期的にPC内のCLを削除する。

この他にも、PC内に該当するprofileが存在しなかった場合にQHへprofile要求する際の待ち行列（Data Request Queues）も上記と同様の仕組みを用いる。

#### (4) 性能評価

同時ユーザアクセス数と応答時間の関係をFP-ProfilerシステムとOracle-Profilerシステムの間で比較し、FP-Profilerの方が応答時間を大幅に低く抑えることができることを実証した。

#### ■ 実験環境

ユーザクリックはArrivalRateを平均した相互到達時間（interarrival time）をもつポアソン過程に従い、NumClicks（ただし、平均MeanClicks、標準偏差StdClicksをもつ正規分布に従う）回の要求を送信した後に終了するようにモデル化して実験する。提供されたナビゲーションリンクのうち何を選ぶかはZip則に従う。同時アクセス数を10, 50, 75, 100（それぞれは5回の平均値）と増加させていき、FP-ProfilerとOracle-Profilerでの応答時間の違いを比較する

#### ■ 評価尺度

評価には、ユーザのクリックからeGlue Serverが応答を返すまでの平均時間

(Average Response Time :  $ART = \frac{\sum (t_{response} - t_{click})}{TotalClicks}$ ) を用いる。

## ■ 実験結果

一般的に、同時アクセス数と応答時間は指数関数的な関係にある。実験ではどちらのシステムにもその傾向が見られたが、FPを用いたシステムの方が増加率をはるかに低く抑えることができた。

表4. 4. 3-3 同時アクセス数による平均応答時間の推移（単位は秒）

	10	50	75	100
FP-Profiler	1	5	10	18
Oracle 8iProfiler	18	55	90	130

[佐藤 祐介]

### 4. 4. 4 Red Opal: Product-Feature Scoring from Reviews

Christopher Scaffidi<sup>1</sup>, Kevin Bierhoff, Eric Chang, Mikhael Felker, Herman Ng, Chun Jin

ACM Conference on Electronic Commerce 2007 pp.182-191

ショッピングサイトにおいてユーザが所望の商品にすばやくたどり着くことを可能にするRed Opalという新しい検索システムについて述べる。このシステムでは、カスタマーレビューから商品カテゴリごとの特徴語を自動抽出し、抽出した特徴語ごとに各商品のスコア付けをする。特徴語抽出の適合率、商品スコアの適合率、特徴語抽出の処理性能、検索時間短縮の推測という4種類の観点で評価した結果、従来システムよりも良好な結果を得た。

#### (1) はじめに

Amazonなどのショッピングサイトで欲しい商品を探すのには手間がかかる。たとえば孫にプレゼントするための「幽霊の本」で検索すると16620件もヒットする。ほとんどのオンラインショップでは、商品の選択には10分から15分を要する。主な理由は、ユーザが大量の検索結果に圧倒されるためである。これに対して、ユーザが満足する商品へと迅速に導くための様々な推薦システムが研究されている。過去のほとんどのシステムは、他ユーザの商品レビューを利用する。

Red Opalはこれらと類似しているが、2つの点が新しい。まず、英語の基本的な単語統計量と確率に基づくヒューリスティックを用いて、商品カテゴリごとの特徴語を精度よく同定する点である。次に、カテゴリに属する商品を、各特徴語についてスコア付けする点である。その結果、ユーザは、ある商品カテゴリ（例：フィクション）を選択し、その商品カテゴリの特徴語（例：幽霊の本）で高い評価を得ているものを迅速に検索することができる。

## (2) 関連研究

特徴語抽出に関しては、Huらの研究<sup>5)</sup><sup>6)</sup>とPopescuらの研究<sup>7)</sup>類似研究である。いずれもタガーにより品詞を付与して、名詞および名詞句を抽出する。HuらはCBAというマイニングツールを用いて、特徴語となり得る頻出名詞句を抽出する。Popescuらは、商品カテゴリと関連する部分全体関係の弁別子と名詞句との間の相互情報量を計算することによって特徴語を抽出する。

単語共起関係を用いる従来手法と異なり、本研究では言語モデルに基づく手法をとる。これは、特徴語については一般的な英語よりも商品レビューにおける使用頻度が高いという仮定に基づいている。一般的な単語頻度統計を用いて専門語を抽出する研究はあるが、商品レビューへの適用例はなさそうである。

## (3) Red Opal

本システムの入力は商品レビューのデータベースである。各レビューは商品の特徴について記述されたテキストと商品に対する評価点からなる。各商品にはユニークな商品カテゴリが付与されている。本システムは、(a)特徴語抽出、(b)商品スコア付け、(c)ユーザインタフェースで構成される。

### (a) 特徴語抽出

- ・基準とする一般的な英語に対する統計量として、100万語からなる英会話および書き言葉のコーパス (British National Corpus) から名詞 $x$ の出現確率 $p_x$ を求めておく。
- ・商品レビューにおける各名詞の出現頻度 $n_x$ と全名詞の出現総数 $N (= \sum n_x)$ を求める。
- ・上記で求めた値と下記の式1により、連続する $N$ 語における名詞 $x$ の出現確率を計算する。

$$\ln(P(n_x)) \approx (n_x - p_x N) - n_x \ln\left(\frac{n_x}{p_x N}\right) - \frac{\ln(n_x)}{2} \quad \dots (式1)$$

- ・特徴語については一般的な英語よりも商品レビューにおける使用頻度が高いという仮定によれば、式1により得られる確率は非常に小さな値となるはずである。実際、デジタルカメラの商品レビュー中に出現する名詞58543語中に「レンズ」という語が1174回出現しており、上式の値は-6429となる。これは非常に小さな数値であり、本システムは「レンズ」がデジタルカメラに関する特徴語であると判定する。
- ・基準とするコーパス中に出現しなかった名詞については、全名詞の平均出現確率を用いて計算する。たとえば「メガピクセル」という名詞について計算すると、上式の値は-488という比較的小さな値となり特徴語であることが推測できる。

### (b) 商品レビューを用いたスコア付け

- ・商品レビュー中のユーザの評価 (星1~5など) と前述の特徴語の出現頻度を用いて、各商品に対して特徴語ごとのスコア付けを行う。具体的には、下記の (式2) により、各レビュー中に出現

する特徴語の頻度に応じて、ユーザ評価に対する重み付けを行う。

$$w(r, f) = \sum_{i=0}^{o(r, f)-1} 2^{-i} = 2 - 2^{1-o(r, f)} \quad \dots (式2)$$

ここで、 $w(r, f)$  は、特徴語 $f$ に関するレビュー $r$ の重みで、 $o(r, f)$  はレビュー $r$ における特徴語 $f$ の出現回数である。特徴語が出現しなければ重み0、特徴語が1回出現すれば重み1、特徴語が2回以上出現すると重みが2に漸近するという式としている。

- 上記の (式2) で計算した重みをもとにして、(式3) により商品スコア $s(p, f)$  を求める。

$$s(p, f) = \frac{\sum_r w(r, f) \cdot rating(r)}{\sum_r w(r, f)} \quad \dots (式3)$$

このスコア付けにはアドホックな点があり、改良の余地がある。

- さらに、各商品におけるスコア分布の標準誤差を用いて信頼性を表すメタスコアを計算する。

$$confidence(p, f) = 1 - \frac{\sigma_w(p, f)}{2 \sqrt{\sum_r w(r, f)}}$$

#### (c) ユーザインタフェース

上記の計算により、商品カテゴリごとの特徴語リストを求める。また、各商品には各特徴語ごとのスコア $s(p, f)$ を求める。本システムのユーザインタフェースでは、これらを直感的に提示するようにしている。

- 本システムのデモ用サイト (<http://redopal.ntelligentsolutions.net>) には、Amazonからダウンロードした8カテゴリ、700製品、5000レビューが格納されている。
- ユーザは、まず商品カテゴリをプルダウンメニューで選択する。(例：PC Games)
- 商品カテゴリを選択したら、次は検索の観点として使用したい特徴語をリスト選択する。(例：expantion packs)
- 特徴語を選択してGOボタンをクリックすると、Red Opalは検索結果の一覧を表示する。検索結果は各製品の名称、スコア、信頼度からなる。スコアは数値で、信頼度は棒グラフで表示される。一覧中の各製品は、Amazonサイト中にある該当製品の詳細説明ページへのリンクとなっており、ユーザが一覧中のいずれかの製品をクリックすると詳細説明ページが表示される。

#### (4) 評価

- Red Opalを以下の4種類の観点で評価した。

(a) 特徴語抽出の適合率： エンドユーザ評価で適合率88% (従来システムは64%)

- (b) 特徴語抽出の性能： レビュー数に対して線形の時間で処理可能
- (c) 商品スコアの適合率： スコアの高い商品に対して適合率80%
- (d) 検索時間短縮： 平均検索時間10~15分を3分に短縮

#### (5) 考察および今後の課題

主な検討課題として、以下の4点を考えている。

- ・ 提供者側の商品説明の記述を利用すること
- ・ 意見語 (opinion-words) の活用
- ・ 特徴語をリストからの選択ではなく自由入力にすること
- ・ 複数の特徴語で商品を検索できるようにすること

[相川 勇之]

#### 4.4.5 Optimal Pricing with Recommender Systems

Dirk Bergemann and Deran Ozmen

ACM Conference on Electronic Commerce 2006 pp.43-51

##### (1) 概要

オンラインマーケットの推薦システムにおける最適な価格設定について報告している。推薦システムは2つの方法でマーケットに影響を与えている：(i) 顧客に対して商品の不確実さを減らすことにより価値を作る、(ii) アドオンとして提供される推薦情報が情報の客観性を生成する。本論文では、推薦システムを使わない競争的周縁 (competitive fringe) に対し、推薦システムを使った売り手の最適価格設定における上記要因の影響を報告している。

また、市場の分割についても報告している。本論文のモデルでは、消費者は、商品に対する嗜好と、嗜好の強さにより分割される。売り手が消費者を、商品に対する嗜好、または嗜好の強さのどちらによって分割するかは、推薦システムのパフォーマンスに依存し、パフォーマンスが悪いときは前者が起り、パフォーマンスが良い場合は後者が起こることを報告している。以下に本論文の想定モデルと市場の分割について説明する。

##### (2) 想定モデル

推薦システムを使う売り手と推薦システムなしの競争的周縁が価格を競い、2段階で商品選択を行うモデルを使用する。段階0では、売り手から異なる2つの商品が提示され、消費者は購入する商品と、購入先とを選択する。推薦システムを使用している売り手は、購入消費者から購入した商品に関する

情報を収集する。次に、段階1で、売り手は消費者に段階0で購入しなかった商品と新たな商品と、段階0にて収集した情報を推薦情報として提示する。そして消費者は、提示されている商品のどちらかを購入する。

このモデルにおいて、消費者の効用値と推薦システムを使う売り手と競争的周縁との関係を示している。

### (3) 均衡（市場の分割について）

完全バイズ均衡において、上記モデルの推薦システムのパフォーマンス（ $\rho$ （値が高い程、パフォーマンスが高い））に対し、5種類の市場の状態を導いている。

状態1： $\rho < \rho_1$ の場合、消費者の商品に対する嗜好によって、購入先（推薦システムを使う売り手 or 競争的周縁）と購入する商品が変わる。

状態2： $\rho_1 \leq \rho < \rho_3$ の場合、推薦システムを使う売り手の市場占有率は100%で、消費者の商品に対する嗜好によって購入する商品が変わる。

状態3： $\rho_3 \leq \rho < \rho_4$ の場合、推薦システムを使う売り手の市場占有率は100%で、消費者の嗜好の流動性によって購入する商品が変わる。

状態4： $\rho \geq \rho_4$ の場合、推薦システムを使う売り手の市場占有率は100%でなくなり、消費者の嗜好の流動性によって購入する商品が変わる。

状態5： $\rho \rightarrow \infty$ の場合、消費者と商品の分布は、完全に混ざり合う。

### (4) 結論

本論文では、上記モデルにおいて、消費者の効用値と推薦システムを使う売り手と競争的周縁との関係を示した。また、推薦システムのパフォーマンスによる市場の分割の状態について述べた。

[續木 貴史]

#### 4.4.6 Recommender Systems and their Impact on Sales Diversity

Daniel Fleder and Kartik Hosanagar

ACM Conference on Electronic Commerce 2007 pp.192-199

##### (1) はじめに

本論文は推薦システムが消費者に与える影響を、顧客の購買の多様性を促進するのかそれとも縮小するのかとの観点に基づき調べたものである。

本論文において、推薦システムが購買の多様性に与える影響は、「新しい商品を発見できるので多様

化する」という考えと「売れる商品がより推薦されるようになるため購買が集中する」という考えの二つに分類されており、推薦システムがそのどちらの性格をもつか、モデリングによる解析とシミュレーションの双方から検証されている。以下にその内容を紹介する。

### (2) 購買の多様性について

本論文でいわれる「購買の多様性」とは、商品の品揃えに対して消費者の購買の幅がどの程度集中していたかを示すものであり、この多様性を示す指標としてジニ係数を利用する。ジニ係数は

$G = 1 - 2 \int_0^1 L(u) f(u) du$  で表される。(  $L(u)$  は一定期間に低い  $100u\%$  の製品が売れた場合の収入である) そ

して  $G_0$  を推薦システムを使わない場合のジニ係数、推薦システム  $r_i$  を使う時のジニ係数を  $G_i$  とし、  $G_i$  が  $G_0$  を上回る場合、下回る場合、同じ値を取る場合をそれぞれ推薦システム  $r_i$  が集中の傾向、多様化の傾向、どちらの傾向もなしと考えることとする。

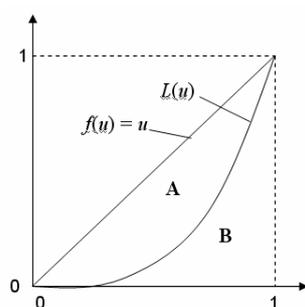


図4.4.6-1 ローレンツカーブ

図中の  $L(u)$  がローレンツカーブを表し、ジニ計数は図中の  $A/(A+B)$  で表される。

### (3) モデルによる解析

本論文では利用するのはTwo-Urn Modelと呼ばれる手法が採用される。この手法では2つのつぼ(Urn)が想定されUrn1には商品(単純化のため白と黒の2種類のみとする)が入っており、確率  $p$  で白が取り出される。この  $p$  の値は顧客が推薦システムなしに白を選択する確率に相当する。一方Urn2に入っている商品は今までの購買履歴が反映されており、関数  $g(X)$  に従って白が推薦される。ここで  $X_t$  は時刻  $t$  より前に白が選ばれた割合であり、 $g$  にはシグモイド関数が採用される。顧客は確率  $r$  で推薦を受け入れるものとされる。

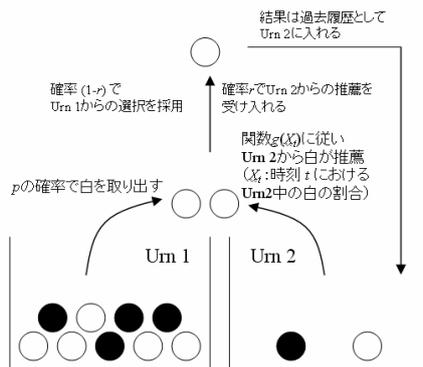


図4. 4. 6-2 two-urn modelの概念図

2つのつぼはそれぞれユーザの意図とシステムの推薦を表し、どちらを採用するかで推薦を受け入れるかをモデル化する。

このモデルを利用し、 $r$ と $p$ の値を変化させた時のジニ係数の値を調べた結果は、サンプルの取り方によって購買の多様性を増やす方向にも減らす方向にも働きうるということである。ただし、多くの $r$ と $p$ の組み合わせについて、 $G_0$ に比べ上昇する、すなわち購買を集中化させる方向に推薦システムが働く。このことから実際の推薦システムの効果を調べるのには一つの販売者について調べるのではなく、同一の推薦システムを使う複数の販売者について、その平均をとって調べる必要があるとの考察が紹介されている。

#### (4) シミュレーション

シミュレーションにより実際に推薦システムを使用した場合とそうでない場合の購買の比較が行われた。ここで用いられる推薦システムは通常の協調フィルタリング ( $r_1$ ) と重要度の違いを排除した協調フィルタリング ( $r_2$ ) の2種類である。後者は「著名な商品は推薦されるべきでない」との考えを取り入れ、計算の際に販売量で割っている。また顧客の選択は多項ロジットモデルを用いて表す。

シミュレーション結果としては、 $r_1$ 、 $r_2$ それぞれに対するジニ計数を $G_1, G_2$ とすると $G_1 > G_2 > G_0$ となることであり、ここで驚くべき点は $r_2$ が購買の多様性を低める方向に働いたことである。

#### (5) おわりに

本論文は推薦システムが消費者の購買に与える影響を調べたものであるが、平均的に見れば協調フィルタリング法が購買商品が集中する方向に働くが、条件次第では購買の多様化を促進するとの結果が示されている。シミュレーションでは、人気の高い商品が推薦されやすい傾向を排除したにもかかわらず

ならず推薦システムが購買商品が集中する方向に働いたことが確認され、また多様化するか集中するかの傾向はデザインにも左右されるとの結果が示されている。

[中臣 政司]

#### 4. 4. 7 Recommending or Persuading? The Impact of a Shopping Agent's Algorithm on Use Behavior

Gerald Haubl and Kyle B. Murray

ACM Conference on Electronic Commerce 2001 PP.163-170

##### (1) はじめに

本論文は、ネットショッピングにおいて、商品推薦エージェントが消費者の購買行動に与える影響を調査した実験について述べている。推薦エージェントは、消費者の好みを把握して、パーソナライズした商品ランキングリストを提示する。しかし、消費者はあらかじめ明確な商品選択基準を持っているとは限らず、商品購入の過程で、順次好みを構築していくと考えられる。このことから、推薦エージェント側で提示する情報を操作することにより、消費者の好みに影響を与えることができる可能性がある。本論文では、模擬ネットショップでテントを購入するタスクの実験を行い、この仮説に対する検証を行っている。

##### (2) 実験設定

本実験では、模擬ネットショップにおいて、被験者に16種類の商品（テント）を提示し、その中から必ず1個を選んで購入してもらう。

###### ① 実験データ

1つの商品は、4つの属性（耐久性、布地、重さ、保証期間）で表される。プレ実験の結果から、4属性の中でも耐久性と重さを重要な属性（プライマリ属性）と位置付けて、結果の分析に用いる。属性は、2種類のグループにわけしておく（グループ1：プライマリ属性が耐久性、セカンダリ属性が布地、グループ2：プライマリ属性が重さ、セカンダリ属性が保証期間）。

また、16個の商品で構成される商品リストとして、以下の2種類を用意する。（図4.4.7-1）

- (a) 各商品の特徴（すなわち、各属性ごとの優劣）に一長一短があり、商品選択が難しいリスト  
(e.g.耐久性がベストな商品は重さが悪く、重さがベストな商品は耐久性が悪い)
- (b) 特徴に一長一短が少なく、商品ごとの優劣が一見してわかりやすく、選択しやすいリスト  
(e.g.耐久性も重さも優れている商品がある)

	商品名	耐久性	布地	重さ (Kg)	保証 (年)
1	Coyote	76	2.3 oz ナイロン	3.4	4
2	Adventurer	76	1.9 oz ポリ	3.4	3
3	Sunlight	79	2.3 oz ナイロ	3.4	4
4	Grizzly	79	1.9 oz ポリ	3.4	3
5	Oasis	82	2.3 oz ナイロ	3.5	4
6	Solitude	82	1.9 oz ポリ	3.5	3
7	Summit	85	2.3 oz ナイロン	3.3	4
8	Drifter	85	1.9 oz ポリ	3.3	3
9	Challenger	88	2.3 oz ナイロン	3.8	4
10	Serenity	88	1.9 oz ポリ	3.8	3
11	Raven	91	2.3 oz ナイロン	3.9	4
12	Waterfall	91	1.9 oz ポリ	3.9	3
13	Naturalist	94	2.3 oz ナイロン	4.0	4
14	Skyline	94	1.9 oz ポリ	4.0	3
15	Neptune	97	2.3 oz ナイロン	3.7	4
16	Freestyle	97	1.9 oz ポリ	3.7	3

重さはこれがベストだが、耐久性は悪い

耐久性はこれがベストだが、重さは悪い

	商品名	耐久性	布地	重さ (Kg)	保証 (年)
1	Traveler	76	2.3 oz ナイロン	3.9	4
2	Journey	76	1.9 oz ポリ	3.9	3
3	Seabreeze	79	2.3 oz ナイロン	4.0	4
4	Moonscape	79	1.9 oz ポリ	4.0	3
5	Galaxy	82	2.3 oz ナイロン	3.5	4
6	Lakeside	82	1.9 oz ポリ	3.5	3
7	BackTrail	85	2.3 oz ナイロン	3.6	4
8	Eagle	85	1.9 oz ポリ	3.6	3
9	Eclipse	88	2.3 oz ナイロン	3.7	4
10	Daydream	88	1.9 oz ポリ	3.7	3
11	Spirit	91	2.3 oz ナイロン	3.8	4
12	Westwind	91	1.9 oz ポリ	3.8	3
13	Glacier	94	2.3 oz ナイロン	3.3	4
14	Wanderer	94	1.9 oz ポリ	3.3	3
15	Mountain	97	2.3 oz ナイロン	3.4	4
16	Outfitter	97	1.9 oz ポリ	3.4	3

どちらでも比較的良好い

図4.4.7-1 2種類の商品リスト（左：商品選択が難しいリストi、右：商品選択しやすいリストii）

なお、商品リストは、被験者がどれか1つの商品を選ぶと、耐久性と重さのどちらの属性を重視したかがわかるように設定されている。つまり、耐久性と重さの両方の値がベストな商品は存在しない。また、全商品の価格は均一である。

## ② 評価方法

本実験で評価する内容は以下の2つである。

- ・ エージェントは被験者の商品購入に影響を及ぼすか？

エージェントは商品の一部の属性だけを使って推薦を行う。そして、推薦に用いた属性と、最終的に被験者が購入した商品の属性を比較することで、推薦が被験者の商品購入に影響を及ぼしたかを調査する。推薦に用いる一部の属性とは、①で述べた属性グループ1,2のどちらかである。

- ・ 商品の違いによりエージェントの影響度に差が出るか？

商品の選択が難しい状況とやさしい状況下で、エージェントの影響度が異なるか、商品リストを変えて調査する。

## ③ 実験手順

実験手順（システムの動作）は、以下の1) - 4) のとおりである。

- 1) エージェントが2属性だけを被験者に提示し、被験者にその2属性の間の重みを0-100の値で入力してもらう（図4.4.7-2）
- 2) 1) で取得した重みに基づいて、商品をランキングする（図4.4.7-3）  
推薦結果の画面では、商品名と1) の2属性の情報のみ表示する（残りの2属性は隠れている）
- 3) 各商品の詳細リンクがクリックされると、詳細ページとしてその商品の4つ全部の属性の値を提示する。詳細ページには、商品一覧に戻るリンクと、購入に進むリンクも提示する
- 4) 商品購入ボタンが押されたら実験を終了する

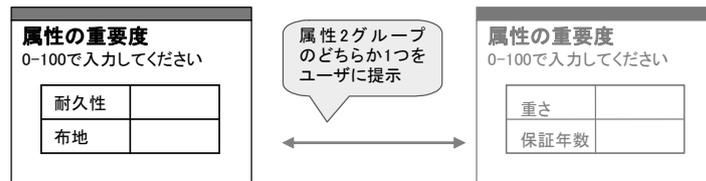


図4. 4. 7-2 属性重要度入力画面（エージェントが推薦に利用する属性をユーザに提示）

最初に被験者に重みを入力させた2属性だけしか見えていない(この2属性を使って推薦)

詳細をクリックすると、4属性全部確認できる

購入は詳細ページから行う

あなた向けのおすすめ商品			
商品名	耐久性	布地	リンク
Neptune	97	2.3 oz ナイロン	詳細
Naturalist	94	2.3 oz ナイロン	詳細
Raven	91	2.3 oz ナイロン	詳細
Challenger	88	2.3 oz ナイロン	詳細
Summit	85	2.3 oz ナイロン	詳細
Freestyle	97	1.9 oz ポリエステル	詳細
Oasis	82	2.3 oz ナイロン	詳細
Skyline	94	1.9 oz ポリエステル	詳細
Sunlight	79	2.3 oz ナイロン	詳細
...	...	...	...

図4. 4. 7-3 商品ランキングの画面

### ③ 実験結果

被験者数347人に対し実験を行い、以下の結果が得られた。

- ・ エージェントは被験者の商品購入に影響を及ぼすか？

60.7%が、エージェント提示の属性を重視し（e.g. 耐久性を提示したら、耐久性が良い商品を購入）、39.3%が、提示と逆の、「詳細」ページに隠れていた属性を重視した（e.g. 耐久性を提示したが、重み重視で商品を購入）。エージェントがないときにどちらの属性を重視するかは50%、50%という仮定からすると、エージェントが消費者の好みに影響を及ぼしたと言える。2項検定の結果、 $p < 0.0001$ で、有意な差が認められた。

- ・ 商品の違いによりエージェントの影響度に差が出るか？

商品特徴に一長一短があり、選択が難しい商品リスト (i) では、71.0%が、エージェントが提示した属性を重視し、29.0%が、提示と逆の属性を重視した（2項検定の結果、 $p < 0.0001$ で有意な差）。一方、商品ごとの優劣が一見してわかりやすく、選択しやすい商品リスト (ii) では、51.5%が、エージェントが提示した属性を重視し、48.5%が、提示と逆の属性を重視した。（2項検定の結果、 $p > 0.75$ で有意な差は見られない）。つまり、商品特徴に一長一短があり、商品選択が難しい場合は、エージェントの影響が大きい。選択がやさしい状況下ではほとんど影響は見られない。

#### ④ まとめ

この論文では、模擬ネットショップにおける購買実験を行い、エージェント推薦と消費者の購買行動に関する相関を調査した。その結果、以下の知見を得た。

- ・エージェントは消費者の商品選択基準に影響を与えることができる。エージェントは商品が持つ属性（重さ、耐久性など）の一部を用いて推薦を行うが、その推薦に使う属性が、商品購入の決定に影響を与えている。
- ・消費者に提示される商品リストによって、エージェントの与える影響度が異なる。すべての属性で優れている商品があり、選択が容易な場合は、エージェントの推薦は購入にさほど大きな影響を与えない。しかし、商品の特徴に一長一短があり、選択が難しい場合は、エージェントの推薦が購入に与える影響が大きい。実際のマーケットは、基本的に商品特徴に一長一短が存在する環境であるので、エージェントの影響は大きいと考えられる。

これらの結果から、推薦エージェントは、消費者の購入意思決定を支援するだけでなく、エージェント側で条件操作することで、消費者の購買行動をコントロールできる可能性も持っていると思われる。

[志賀 聡子]

## 4.5 SIGIR2007参加報告

### 4.5.1 はじめに

2007年7月23～27日にオランダのアムステルダムで開催された「第30回ACM SIGIR年次大会（以下、SIGIR2007）」について報告する。

SIGIR (Special Interest Group on Information Retrieval) は、情報検索に関するACM (Association for Computing Machinery) の研究部会であり、SIGIR2007は30回目の年次大会である。情報爆発時代の到来によって情報検索技術の重要性が高まっている。こうした状況の中、SIGIRの年次大会は回を重ねるごとに規模を拡大している。SIGIR2007では、口頭発表に490件という過去最多の投稿があり、その中から85件の投稿が採択された。17%という採択率は情報処理関連の国際会議としては狭き門であり、情報検索に関する研究開発が国際的に活発化している様子が分かる。参加者は約600名だった。

プログラムは以下の通りであった。

- ・7月23日: チュートリアル (9件)
- ・7月24～26日: 本会議
- ・7月27日: ワークショップ (9件)

本会議では、基調講演、口頭発表、ポスター発表に加えて、Karen Sparck Jones氏によるビデオ講演があった。以下、チュートリアル、本会議、ビデオ講演について紹介する。筆者はワークショップには参加しなかったので報告の対象からは割愛する。

### 4.5.2 チュートリアル

SIGIR2007の参加者は午前と午後から1つずつチュートリアルを選ぶことができた。筆者は、以下のチュートリアルに参加した。

- ・ The Probabilistic Relevance Model: BM25 and beyond

今や情報検索の標準的なモデルとして使われているOkapi BM25について、歴史的経緯や今後の展望などについて解説された。

- ・ Foundations and Challenges of Web Advertising

Web上のビジネスモデルとして急速に発展している「Web広告」について、最先端の研究動向が解説された。本チュートリアルはYahoo! Researchのメンバーによって行われた。同メンバーは

「Computational Advertising」という領域を提唱しており、本会議においてもWeb広告に関する口頭発表を2件行った。

#### 4.5.3 本会議

本会議は、3つの口頭発表セッションが並行して行われた。17%の採択率を勝ち抜いた研究だけあって、聞き応えのある発表が多かった。ほとんどの発表が大規模なテストコレクションや検索ログを用いた実験を行っており、さらに統計検定によって実験結果の妥当性を検証していた。採択率が厳しいため、提案手法によって結果が改善したことを示すだけでは不十分である。結果の改善が必然か偶然かを調べるために統計検定によって有意差を示すことは、半ばSIGIRに採択されるための必須要件となっている。

28あるセッションのうち、本委員会の活動に関連が強いセッションは以下の通りであった。

- Personalization
- Routing and Filtering
- Users and the Web
- Web IR
- Interaction

その他、Learning to Rankというセッションでは、近年の傾向として、情報検索に機械学習を応用した発表が行われた。具体的には、文書を順位付けるためのモデルや関数を学習する手法が発表された。

Web検索に関する研究では、検索ログを用いた研究が盛んであった。検索ログとは、ユーザがWeb検索システムに入力した検索質問と、検索結果のうちクリックされたページ（クリックスルー）に関する履歴情報である。検索ログをマイニングすることで、「キーワードとページの関係」、「キーワードどうしの関係」、「ページどうしの関係」など種々の関係を発見し、検索に応用することが可能になる。

ただし、マイニングに値する大規模な検索ログは検索サービス大手の企業でない限り入手が難しい。SIGIR2007では、Google、Yahoo!、Microsoftのような検索サービス大手か、もしくは企業から提供されたログを利用した大学の発表があった。なお、Googleが研究発表を活発に行うようになった点は印象的だった。

以下に、検索ログを利用した発表をいくつか紹介する。括弧内には発表者の所属機関名を示す。

- Learn from Web Search Logs to Organize Search Results (University of Illinois at Urbana-Champaign)

検索された文書をクラスタリングするために検索ログを利用する。要点は、クラスタリングの「観

点」をログから抽出することと、生成されたクラスタにログ中のキーワードを用いてラベルを付与する点にある。

- Information Re-Retrieval: Repeat Queries in Yahoo's Logs (MIT, University of Washington, Yahoo! Research)

発表者らの調査によれば、検索エンジンに入力される質問の約40%は、過去に調べたことを再び調べるために使われる「再発見質問 (re-finding query)」である。本研究は、再発見質問に対して以前と異なる結果を返すとユーザが目的の情報を見つけづらくなることを明らかにした。また、再発見質問を認識する手法について提案した。

- Studying the Use of Popular Destinations to Enhance Web Search Interaction (Microsoft Research)

ある検索質問を入力したユーザに対して、「過去に同じような質問を入力したユーザがクリックしたページ」を提示する対話的な検索手法を提案した。既存の検索エンジンでは、検索質問と関連する言葉 (関連語) をユーザに提示する対話的な検索手法がある。関連語による検索は既知事項検索 (known-item search) に適しているのに対して、本研究で提案された手法は探検的検索 (exploratory search) に適していると発表者らは主張している。なお、本発表はBest paper awardを受賞した。

- Random Walks on the Click Graph (Microsoft Research Cambridge)

クリックスルーデータは、検索質問と適合文書のペアと見なすことができる。本研究はクリックスルーデータから、ある検索質問に対するページの順位を学習する手法を提案した。しかし、クリックスルーデータでは、ユーザが本当に適合情報と判断してクリックしたかどうかは分からないため、雑音が多い。本研究で提案された手法の特長は、雑音入りのクリックデータに対して頑健であることと、クリックされていないページに対しても順位を付けることが可能な点にある。

#### 4.5.4 Karen Sparck Jones氏のビデオ講演

情報検索研究の先駆者であるKaren Sparck Jones氏は2007年4月に逝去された。本セッションでは、同年2月にSparck Jones氏がACM Athena Awardを受賞した後で撮影された講演のビデオが上映された。ビデオ上映に先立って、Donna Harman氏やStephen Robertson氏によってSparck Jones氏の研究業績や研究にまつわる逸話が紹介された。現在の情報検索において標準的な技術として確立している「IDF (Inverse Document Frequency)」はSparck Jonesの業績である。業績紹介の中で、以下の

論文はIDFの原点として紹介された。

Karen Sparck Jones.

A Statistical Interpretation of Term Specificity and its Application in Retrieval.

Journal of Documentation, Vol.28, No.1, pp.11-21, 1972.

#### 4.5.5 おわりに

本節は、SIGIR2007のチュートリアルと本会議に関する参加報告を行った。2008年の大会は7月20～24日にシンガポールで開催される予定である。

- ・ SIGIR2007のホームページ <http://www.sigir2007.org/>
- ・ SIGIR2008のホームページ <http://www.sigir2008.org/>

[藤井 敦]

#### 参考文献

- 1) 神寫 敏弘, "推薦システムのアルゴリズム(1)", 人工知能学会誌, vol.22, no.6, pp.826-837(2007)
- 2) 神寫 敏弘, "推薦システムのアルゴリズム(2)", 人工知能学会誌, vol.23, no.1, pp.89-103(2008)
- 3) 神寫 敏弘, "推薦システムのアルゴリズム(3)", 人工知能学会誌, vol.23, no.2,(2008)
- 4) J. Ben Schafer, Joseph A. Konstan and John Riedl: "E-Commerce Recommendation Applications", Journal of Data Mining and Knowledge Discovery, Vol.5, No.1/2, pp.115-152(2000).
- 5) M. Hu and B. Liu: "Mining and Summarizing Customer Reviews", Proc. 10th SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining, pp.168-177(2004).
- 6) B. Liu and M. Hu: "Cheng, J. Opinion Observer: Analyzing and Comparing Opinions on the Web", Proc. 14th Intl. Conf. on World Wide Web, pp.342-351(2005).
- 7) A. Popescu and Etzioni: "O. Extracting Product Features and Opinions from Reviews", Proc. Joint Conf. on Human Lang. Tech. / Conf. on Empirical Methods in Natural Lang. Processing, pp.339-346(2005).



—— 禁 無 断 転 載 ——

本報告書に掲載されている会社名および製品名は、各社の登録商標または商標です。注記がない場合もこれを十分尊重します。

### 知識情報処理技術に関する調査研究報告書

発 行 日 平成20年3月

編集・発行 社団法人電子情報技術産業協会

知識情報処理技術専門委員会

インダストリ・システム部 企画グループ

〒101-0065

東京都千代田区西神田3丁目2番地1号

千代田ファーストビル南館

TEL (03) 5275-7261

印 刷 三協印刷株式会社