2010年10月22日発行　第20巻3号　　学術刊行物 ISSN 0917-1436

Journal of Japan Society of Information and Knowledge

情報知識学会誌

Vol.20　No.3　(Oct. 2010)

～～～～～～～～～～～～～～～～～～目　次～～～～～～～～～～～～～～～～～～

～～～

JSIK
情報知識学会
http://www.jsik.jp/

巻頭言

将来の学術の担い手と学会

常務理事　長塚　隆

一昨年に始まった米国発の金融危機以来、我が国でも「派遣切り」や「年越し派遣村」など状況はさらに厳しさを増しているように思われます。昨年秋には、初めての選挙による政権交代が起き、様々な期待の下で新らたな政権がスタートし、マニフェストの実現、政治主導、官僚依存からの脱却など新しい試みが行われています。そのような中で、昨年 11 月には来年度予算に向けての「事業仕分け」が始まりました。

そこでは、当学会にも関係が深い科学技術関連事業や高等教育への施策などに対して、即効性や成果の見えにくいなどの理由で、多くの事業の削減や縮小などの方向性が強調されました。新しい政権のもとで、行政刷新会議が事業仕分けの対象とした文部科学省関係の事業についての仕分け結果については、15 万件以上の意見が文部科学省に寄せられたとのことです。このような動きは、今後の当学会の事業にも様々な影響を及ぼしてくるのではないかと思われます。このような状況の下では、学術研究の持つ社会的な意味について、より多くの人にさらに理解してもらうための企画などが大切になっていると言えるでしょう。

研究者の集まりである学会の多くとも密接な関係にある日本学術会議では、新しい試みの一つとしての事業仕分けについて、昨年 11 月 20 日に、会長談話で、予算編成に当たって「人文・社会科学を含む基礎研究から開発研究に至るまでの学術研究を重視」してほしいこと、科学・技術の成果は一朝一夕に成るものでないことなどを指摘し、もし、基礎研究への投資がたとえ短期間であっても大きく減少するようであれば、研究を実際に担う人材が離散し、その結果として、長期的には国際的な競争力にも影響してくると述べています。

さらに、「日本の未来世代のために我々が今なすべきこと」という日本学術会議の幹事会声明を本年 1 月 15 日に出し、「持続可能な人類社会と日本社会の展望」を切り開くためには、「人文・社会科学から自然科学まですべての学術的活動」の総合力を発揮することが必要であり、現在の「出口としての技術をもっぱら重視する科学技術政策」から「基礎研究をしっかりと位置付ける総合的な学術政策」への転換が必要であるとしています。このように、現代の学術研究は、社会や政治との関わりがますます大きくならざるを得ない環境にあると言ってもよいかも知れません。

特に、昨年の事業仕分けの際に、若手研究者の育成をどのようにしてゆくかということで多くの議論がありました。若手の研究者が成長できる環境をどのように用意できるかは、社会全体だけでなく、それぞれの学会にとっても重要な課題になっています。

情報知識学会においても昨年の総会で、会費の見直しについて検討を進めることが承認され、理事会で具体化することになりました。様々な、意見が出され、検討された結果、新たな制度が公表されました。その中では、経済的に困難な状況にある若手研究者を対象にし、会費を半減した若手向き会員種別「ユース会員」制度などが新設されました。これらの取り組みは全体からみれば小さなことかも知れませんが、少しでも歩みを進めることが大切でしょう。

それぞれの大学や研究機関あるいは民間企業の中で個別では取り組み難い事柄も多くなっている時期に、多様な年齢層の研究者が対等の立場で横断的に参加した組織である学会の役割はさらに大きくなっています。

特集紹介

コンピューターネットワークを用いた新観光サービスの提案と情報危機管理：技術者の実践的な育成法について

和歌山大学　システム工学部
教授　中川　優

今回の招待論文は、コンピューターネットワークにおけるサービスと運用において、これから益々脚光を浴びると思われる２分野から、ごく最近の研究開発の様子を会員の皆様にお伝えしたいと思い選びました。本分野は自分の専門分野でない方は、ざっくりと大まかな内容をつかんでいただき、また、専門の方は他の文献などを参照して今後の研究開発にお役立てください。

一つ目の論文のタイトルは「観光サービスにおけるユビキタスシステムの開発と実証」です。日本の観光事業をヨーロッパのフランス、イタリアに早く追いつけるようにと和歌山県を挙げて観光学科の創設を目指し、国立大学法人では初めて平成１９年度に和歌山大学に開学科し、平成２０年度より観光学部としてより強力に進めてきています。和歌山県は気候が温暖で果物や新鮮な魚などが豊富で、良質な温泉地も多く、各地からの観光客で賑わってきました。　また、「紀伊山地の霊場と参拝道」が平成１６年７月に世界遺産に登録されてからは、日本国内や東南アジアや他の国々からの観光客も増え国際色も豊かになってきました。今回ご紹介する観光案内システムは、平成２１年度総務省ユビキタス特区事業として採択されたものであり、日本語と英語で使えるようになっています。特徴として、行政の防災システム等との連携をとるなど、地域の基幹システムとして発展する可能性が十分あると思います。

二つ目の論文のタイトルは「情報危機管理における演習環境の構築と運用」です。

インターネットの普及により、様々な情報がネットワーク上で管理され、簡単に利用されるようになってきました。それにつれて利用者も多様化してきており、個人情報の漏えい、個人に対する中傷・非難など法律的、道義的責任を問われる事象が急増してきています。しかし、これらの問題を解決するには、高度な技能と幅広い経験が必要となり、簡単には人材の養成が出来ない状況にあります。本論文では、実戦形式により、２０以上の演習事例（シナリオ）を用いて人材を育成します。既に、２００６年より「サイバー犯罪に関する白浜シンポジューム」を毎年開催し、また、「ＩＴkeys-先導的ＩＴスペシャリスト育成推進プログラム」においても、２００８年より毎年実施してきています。前者の白浜シンポジュームでは、２００８年に経済産業大臣賞を受賞し、後者のスペシャリスト育成推進プログラムでは、２０１０年度に文部科学省より「世界最高水準のＩＴ人材育成に向けた成果」との評価を受けました。まだまだ研究途上ではございますが、読者の皆様のお役に立てればと思い、招待論文として執筆していただきました。

招待論文

観光サービスにおけるユビキタスシステムの開発と実証

Development and demonstration of a ubiquitous system for tourist services

下山二郎
Jiro SHIMOYAMA

株式会社見果てぬ夢
The Impossible Dream, Inc.
〒142-0064 東京都品川区旗の台 1-10-7　シャルトルーズセブン C 棟
E-mail:shimoyama@ip-dream.co.jp

近年，オフィスや自宅を離れた屋外でのICTの活用が活発になってきている．その一例として，旅行分野におけるICTの活用が旅行者にとって実際に便利かどうかを検証するために，ユビキタス観光システムを開発し実験した．実証にあたっては，海外からの旅行者を含めた数十人による２回の利用実験を行い，その結果をまとめた．開発したサービスでは，観光ガイドブック等の情報を，携帯電話（アイフォン等を含む）をネットに接続して，静止画，動画，音声ガイド等としてGPSと連動させて閲覧できる他，カメラ機能を活用して写真をマイページにアップロードし，旅行後にも，フォトメモリーや他の旅行者へのお勧めスポットとして利用可能とするものである．結果としては，ガイドブックを携行する煩雑さが避けられ，最新の情報が得られるなど好評であったが，歩行中に閲覧することによる事故の発生等も想定されるなど，今後の研究の課題も残している．

Recently ICT have become to be used not only in offices and homes but also outdoors. We developed a new information system to provide tourists with information on the spots via mobile data networks where the information should include their history and culture in the forms of photos, movies, texts and audio-guides. Tourists can also upload their photos and share them with the others. The experiments have shown the system good and convenient for tourists as they can travel with the latest information without carrying heavy guidebooks. We found some problems such as the risk of accidents in walking caused by concentration on the phone display, and will continue the development to solve those problems.

キーワード：ユビキタス，高速携帯電話NW，ICT路上利用，旅行用ICT，仮想ガイドブック
Keyword：ubiquitous，LTE (Long Term Evolution mobile network)，outdoor ICT applications，ICT for tourism，virtual guidebooks

1. 事業概要

1.1. 事業目的

1.1.1. 本プロジェクトの背景

本実証実験は，言葉の通じない外国を旅する個人旅行者が，通訳兼ガイドを同伴できない場合にも，安全で確実な旅行ができるようなサービスの実験として，平成21年度の総務省ユビキタス特区事業として採択されたものであり，システム開発，モニターツアー等を実施した．

観光における情報（観光資源，歴史，文化，施設等）をユビキタスに活用することで旅行の利便性を飛躍的に高めるため，ユビキタス研究所が提唱している「uコード」を「ユビキタス番号」として用いて，情報を登録，管理，検索するとともに，AR(Augmented Reality，拡張現実)と呼ばれる手法を用いた仮想ガイド機能を携帯端末に提供することによって，着地型旅行会社(デスティネーション・マネジメント会社，DMC）を利用する旅行者に対して英語等による案内サービスを提供しようとするものである．

1.1.2. 事業の目的

和歌山県田辺市を中心とした熊野地域において，外国人個人旅行者向けのDMCに旅行商品管理システムを導入し，情報発信受付窓口を一本化するとともに，WEB及びインフォメーションセンターでの旅行の申込・受付・決済のワンストップサービスを提供する．

AR技術は，現実の環境(風景等)に対して，緯度経度，時間等を元にデータベースから得た他の情報(過去の情報，記録者の情報等)を重ねて記録・照会する技術であり，携帯電話，PC等のデバイスやデジタル写真等からWEB経由でセンターにアクセスし，データの照会を可能とするものである．

外国人旅行者に対して，携帯端末（レンタル携帯電話）や観光ポイントに設置した固定端末を通じて，対面コンシェルジュ，自動翻訳，ナビゲーション，予約，斡旋，照会等のサービスを提供する．これにより，旅行者は申込時から，旅行シーンに応じたWEB，通話，対面での英語対応サービスを受けることができる．

1.2. 事業内容

1.2.1 ICTを活用した新しいサービスモデルの確立

(実施内容)

(1)技術開発・システム構築

ア.ユビキタスサービス要件に関する調査・検討－個人旅行者に対する標準化表示等の整理

イ.システム開発，プログラミング－ユビキタスシステムの開発

ウ.実証用システムの構築－設計，開発，テスト

(2)サービスの実証

ア．実証の準備

実施体制の構築，参加者の募集，機器の設置，風景・施設・バス停留所のイメージデータ収集

イ．フィールド実証の実施

モニターツアーを実施し，モニター参加者(30～50名程度)に対して，携帯電話による仮想ガイド，名所等の仮想掲示板，表示の標準化，マイマップの作成及び旅行予約情報との連動について実証実験を行う．また，宿泊施設，プログラムサプライヤー，DMC管理システム及びDMCユビキタスシステムの実証実験を行う．

ウ．実証データの取得・分析

アンケートの設計・回収，実証データの取得・分析・仮説の検証，課題抽出

(3)ビジネスモデルの検証

ARのユーザビリティ，実用性，体系化

されたユビキタス番号の付与による観光情報のデファクトスタンダードの確立，周辺の飲食店や施設との連動による市場創造の可能性の検証，デバイスの有用性ならびに可能性について検証する．

(4)標準技術の確立

実証実験に基づき，ユビキタスコードと旅行情報に関する総合的な体系に基づく各種技術評価(ユーザ評価，旅行関係者評価，サービスオペレータ評価等をアンケート等により実施)を行う．

アンケート等及び実証実験分析に基づき最適な技術標準を実証項目毎に定義し確立する．

(5)制度等の確立

実証実験における上記アンケート及び実験データの分析に基づき，旅行サービスにおけるユビキタスコードの活用に関する取扱規定等を作成し，今後の利用を促進するためのルールを確立する．

2. ICTを活用した新しいサービスモデルの確立

2.1.ユビキタスシステムの技術的概要

実現しようとするサービス，ビジネスの技術的特色は以下のように要約できる．

①海外からの旅行者やシニアの旅行者が観光を目的に現地を訪れた際に，ガイドブック等の情報だけでなく，最新の情報や旅行情報(予約情報等)を現地でスムーズに得られ，あるいは活用できるサービスを，ICT技術を応用して提供する．

②通信技術としては，携帯電話を中心とするインターネットへの接続とアクセスしたWEBポータルとの連動サービスを統合的に活用する．(固定電話も含む)

③最先端のバーチャルリアルティ技術(旅行現地等において，現実の景観にDBからの情報を表示すること等)を活用する．

④本サービスで中心的に活用する統一化された観光関連「ユビキタス番号」(以下UQ番号※)を自動生成する機能とその番号に対応する情報整理をする機能を提供する．

⑤UQ番号に紐付けられてDB化された情報を，利用者がUQ番号に関連付けられたコードにより携帯電話，PC等で検索・表示する機能を提供する．

⑥これらの情報を旅行者が再利用できるように，個別の記録(マイ旅行ページ)を作成する機能も合わせて提供する．その際に表示する言語，情報等を選択・活用する機能も提供する．

⑦情報の収集が必要な場面（史跡，美術館，宿，空港，食堂等）において，旅行者がDMCのUQセンターにインターネットアクセスすることにより，必要な情報（歴史情報，観光情報，宿泊予約情報，交通情報，飲食情報（食材情報等）をUQ番号に関連付けられたコードを利用して検索することにより，簡便に最新の情報を，多言語で提供する．

⑧旅行関連事業者(宿泊施設，飲食施設，美術館等施設，空港等交通施設他)が，これらの情報を旅行者と確認しあうことで確実な対応及び多言語への対応が可能となるような機能を提供する．

⑨旅行事業者等にDBへの登録機能，DMCセンタースタッフには登録及び管理機能を提供する．登録は携帯電話等からも簡単に誰でもできる機能を実装する．

⑩旅行者に，旅行の写真や口コミ情報等を個人ページや総合ページに登録できる機能を提供する．これに対して他の旅行者や旅行事業者が生の声，意見等を付加，閲覧できる機能を提供する．

⑪旅行者が旅行前に着地型旅行を計画するためのプランニング機能，予約機能，及び決済等機能を提供する．(UQ番号と連

動することで利用がスムーズとなる機能）

⑫旅行者のプラン，予約等に沿ってDMCスタッフが現地手配及び決済，管理等ができる機能を提供する．(UQ番号化されることで管理・利用をスムーズに実現する機能）

⑬地方自治体等には，旅行者が毎日どのように行動し，情報を取得活用しているかを個人情報を削除した一般情報の形で提供する．（レポート機能）

⑭地方自治体等で必要な緊急情報等を登録，管理する機能を提供する．この機能は旅行事業者及び旅行者に最新の行政情報（安否確認等を含む)を提供する．

⑮地域経済の中心となる飲食店，物産店等に，UQ番号と関連付けられたコードを通じて利用，申込，購入，決済等が出来る機能を提供する．

※「UQ番号」(UQコード)は本実証実験で各観光情報に対して生成される番号で，「uコード（主体）＋uコード（客体）＋uコード（関連性）」という構成である．uコード（ユビキタスコード)はUidセンターで作成される主体コード，客体コード，関連性コードである．uコードの発番はTエンジンフォーラムの運営するuコードサーバから，利用者(本実証実験ではUQシステムサーバ)がネットワーク経由で取得し，これによりuコードの二重利用を防ぐ．

本実証実験では，利用者が128ビットの番号を手入力するのが不可能なため，あらかじめ検索用にジャンル分けしたコンテンツに，View NO，歴史NO等として3桁の数字のUQ番号を付与し，サーバ内で関連づけを行う．これを，本事業では「UQ番号」と呼ぶ．これは2次元バーコード(QRコード)で提供することも可能であり，またGPS位置情報やAR技術による空間認識とリンクさせて読み取ることも今後の課題としている．

3. サービスの実証

3.1.実証実験の流れ

携帯電話による観光情報システムの実証

実験URL：

http://kumano.ictu.ip-dream.jp/icts/

(1)情報検索

①ログイン－実証実験共通アカウントを入力して「ICT観光ユビキタス実証実験熊野古道モバイルサービス」にログインする．

②言語切替－日本語と英語の言語切替えを行う．

③宿泊・歴史・食事・交通情報の検索－メニュー画面でカテゴリー(全体，歴史，宿泊，食事，交通)を選択し，検索する．カテゴリーがわからない場合は「全体」から検索する．

④情報検索－知りたい情報の名称か，あらかじめ割り当てられたView NOを情報検索画面で入力する．または「緯度経度ボタン」を押すことで，検索カテゴリーにあった周辺の登録ポイントを検索する．

⑤詳細情報画面－関連情報がある場合は詳細画面からジャンプできる．詳細画面についての音声案内がある場合は，音声再生が可能．

(2)マイビューポイント・マイコメントの登録

マイコメント機能を用いて，熊野古道の各ポイントで，一息つきながら携帯に写真やコメントを登録できる．熊野古道ウォーキングの足跡を残し，思い出の旅記録を作ることができる．他の旅行者のコメントも閲覧，参考にできる．

①携帯で写真や動画を撮る．

②メニュー画面「マイビューポイント」からアクセス.

③「緯度経度送信」ボタンを押して登録される位置を設定すると，その場所に記録される.

④コメントを自由に記入.

⑤「気持ち」をマーク.

⑥「カテゴリー」分けも可能.

⑦「公開」を選択すると，その場所で他の人がアクセスした際にそのコメントを読むことかできる.

⑧携帯で撮った写真や動画もガイダンスに従い操作すると登録でき，コメント閲覧時に画像も見ることができる.

(3)マイコメントの閲覧

熊野古道の各ポイントで，他の旅行者が登録した写真やコメントを閲覧できる.閲覧の足跡も確認できる.

①メニュー画面の「マイビューポイント」からアクセス.

②閲賢したい場所を設定.

閲覧は，つぎの２つのエリアで登録されている公開コメントを見ることができる.

ア)現在いる場所近辺に登録された情報⇒「緯度経度送信」ボタンを押す.

イ)熊野古道View NO（近辺で登録された情報⇒閲覧したい名称かView NO(例:114，116など，資料参照)を入力し，「検索」ボタンを押す.

マイコメントでは対象エリアで登録された全ての「公開情報」を閲覧することができる.これらは地図上に表示したり，時間順に並べて見ることもできる．対応携帯電話機種では動画も見ることができる.

3.2.モニターツアーの実施

携帯電話による情報提供の実証実験を行うため，モニターツアーを実施した.

(1)平成22年2月6日（土）11:30～15:00，曇り雪．ＪＲ紀伊田辺駅～道の駅中小路牛馬童子ふれあいパーキング～近露王子～野中の一方杉．参加者８名（内外国人１名）.

(2)平成22年2月27日(土)11:30～16:00，曇り．ＪＲ紀伊田辺駅～道の駅中小路牛馬童子ふれあいパーキング～近露王子～野中の一方杉．参加者１７名（内外国人６名）.

3.3.ビジネスモデルの検証，システム・サービスの検証

3.3.1.ユビキタスシステム評価

(1)システム全体の概念

ユビキタス(UQ)システムの要点は，下記の4点にあると考え，これにあわせたシステム開発と実証実験を行った.

ア．UQシステムの稼働

旅行前の計画(口コミ情報，景色情報，交通情報等)の作成及び準備(洋服，カメラ等)から活用が開始される.旅行地に到着後は，携帯電話を中心とした利用となる.

イ．UQシステムに蓄積される記録の活用

観光施設の予約，利用，またこれらへのコメント，歴史的景観，観光景観等への足跡の記録，コメント等を，個人情報保護を前提として活用し，マーケティングや観光サービスの商品化，強化，発展につなげる.

ウ．UQシステムのコミュニティ的活用

上記の個人利用履歴やコメント等を，他の旅行者が有益な情報として活用することで，着地型観光事業を向上させ，観光地の持続性の高い発展に寄与する.

エ．UQシステムの行政サイドの活用

行政サイドにおいては，国際観光地としての地位向上に不可欠である安全，安心，即応の視点での活用が可能である.

例えば防災情報等を観光UQシステムと連動させることで，観光地で自然災害等に遭遇した場合，そのエリアにいると思われる旅行者の把握(言語,行程,その他)が可能となる．また，それぞれの所持する携帯電話に対して，適切な防災関連情報をリアルタイムに言語属性に応じて提供することで，旅行者の安否確認や安全を確保することが可能となる．その際に，従来の標識やテレビ，ラジオ，防災無線等に加えて，個人の属性に合わせた防災情報をユビキタス情報として提供し得る．

(2)ユビキタスシステムの評価

上記4点に対する現状でのシステム評価は，以下の通りである．

ア．UQシステムの稼動

UQシステムの稼動に関しては，AR技術の活用に関しても，電波環境の影響を受けることが多いが，交通情報等は電波環境の良い場所での利用が多い，宿泊施設，観光施設等も同様に電波環境が良い場所にあることが多いことなどにより，一定の評価を与えることができる．また，熊野古道のように山奥の環境でも，GPSにより携帯電話に蓄積された周辺情報の表示は可能である．

しかし，電波環境の悪いところでは情報取得に時間がかかる．一般的には2秒以内で設計しているが，実際には10秒程度かかる場合もあった．この点に関してシステム全体の評価は低いと考える．

今後，携帯電話に蓄積できる情報量の増大や機能の強化が進むことによって，携帯電話等のユビキタス活用がより効果的になると考えられる．

例えば，電波等が受信しにくい環境では，UQシステムの効果的な活用には，データセンターから逐次送り込まれる情報を，携帯電話の内部メモリーやSDカードに蓄積して活用できるようなシステムであることが，重要と考えられる．

イ．UQシステムに蓄積されてゆく記録の活用(③)

UQシステムに蓄積される記録としては，宿泊施設，観光名所，歴史情報など，観光事業者サイドが用意するコンテンツを，旅行者が活用することがまず考えられる．今回のシステムでは，概ね計画通りで，和室の利用方法や温泉の入り方等がわかりやすいという点など，外国人利用者からも評価が高いものとなっている．

また，これらの情報はユビキタス番号によりジャンル区分に従い体系化されており，他の宿泊施設や観光名所等でも活用可能であるため，今後評価が高まるものと考えられる．

一方，観光客自身がメモ的に記録する「感動」，「日記」，「おいしさ」等の情報は,GPS情報との連携によって付加価値が高まるシステムとなっていることから，利用者から高い評価を受けた．他の旅行客が登録した情報は，口コミ的に現地で活用・展開可能なシステムであり，この点で評価は高かった．

今後の課題としては，これら蓄積情報関してよりわかりやすく，使いやすいものとする必要があり，この点に関して本システムは平均レベルの評価であると考えている．

ウ.UQシステムのコミュニティ的活用

DMCがめざす着地型観光事業は,大量送客タイプの観光事業と違い，観光地と利用者がより深くつながるという特色をもつと考えられる．この深いつながりはさらに，旅行者同士のコミュニティ化を進め，相互に緩やかな連携を創造できる可能性をもつものである．例えば，すれ違う際に声を掛け合う，すばらしい景色を教えあうというコミュニケーションがリアル(同時的・共時的)に行われ，コミュ

ニケーションの質の高い観光地となることで，さらに集客とが可能となる．これらのコミュニケーションの支援ツールとして，UQシステムは，リアル，ノンリアル(同時に旅行していなくとも)に情報やコミュニケーションを新たな観光資産として創造することを可能とする．この点で非常に評価の高いシステムとなっている．

例えば，春の旅行者が，秋の旅行者とノンリアルに情報を交換することで，それぞれの観光地に対する造詣や信頼，絆が高まるといった効果があると考えられる．

これらのことにより，質が高く，継続性が高い観光事業の創出にUQシステムは価値があり，支援システムとして評価が高まると考えられる．

エ．UQシステムの行政サイドでの活用

行政サイドのシステムとしては，さまざまな利用が可能であるが，今回の実証実験においては，防災という観点からも評価が高いものであることが認められる．実証実験中に降雨，降雪があり，交通ダイヤ等の乱れを，本システムを通じて旅行者が確認することができた．

また,実際に2月27日のチリ地震における津波情報等も一部配信したが，これは海外からの旅行者に評価された．

4.観光情報の国際標準化に向けたUQコード活用の提案

(1)UQコードの活用

上述のように，今回の実証実験において開発したユビキタスシステムに関して以下の点よりUQコードの活用を提案する．

ア．ユニークなコードとしての分類や蓄積，利用の容易性

観光という幅広い情報を，上述のようにシンプルに定義しても，継続的な観光事業等への活性化するためには，無限大に近いコードをベースとするコード技術を活用すべきである．

この点においてuコードを主体，客体，関連性の3コードをベースとして，1つのUQコードを生成するという方式を提案する．

イ．連携性に関する有用性

既に存在する観光システム，旅行情報システム等に，今回の実証実験で開発をしたUQコードシステムを適用し，それぞれの持つ個別，単独なコード体系にUQコードを重ねて付与することが可能である．これにより観光地ごとに独立していた情報を連携して利用できることになり，幅広い観光市場の情報管理，マーケティングへの活用が可能となる．

ウ．観光以外のサービスとの連携による発展性

UQコードは，既にさまざまな分野で活用が始まっているuコードをベースに構成されており，将来的に，カーナビゲーション，食育，農業等々のuコードと連動させ,「農業観光」，「食観光」等々への展開が可能となるので，この点からもUQコードの活用を提案する．

上記の諸点で,uコードを用いて構成される観光情報を管理する個別コードであるUQコードは優れており，今後，日本を始め，アジア，欧米の観光地等へも利用拡大を提言，提案していきたい．

5.まとめ

本稿では，昨年度にICT経済・地域活性化基盤確立事業（ユビキタス特区事業）に採択された「観光情報の国際標準化を目指すためのユビキタス実証実験～誰もが使える体系化されたユビキタス番号」

（提案代表会社(株)見果てぬ夢，プロジェクト責任者下山二郎）において行われた実証実験の中から，移動中の観光者に対するユビキタス技術の有用性の実証に関する部分を抽出し記述した．

本システムは現在実商用化に向けて着々と準備を進めている．今や携帯端末やスマートフォン等の発展により観光ガイドブックのモバイル化が一層進んでいる．この関連の端末アプリケーションにはゲーム感覚を優先し，旅行そのものよりゲームを楽しむという観光振興的でないものも見受けられる．また，来年度以降に登場する1メートル単位で測位可能なGPSや，高速携帯ネットワーク，携帯TV，タッチパネル端末等の出現，発展により，観光事業はさらに有望な市場となることが予想される．こうした状況を踏まえつつ，今後さらに研究を進める必要があろう．

謝辞

本実証実験には総務省を始めとして様々な団体，組織，研究者のご支援を頂いた．ここに記して御礼申し上げる．

招待論文

情報危機管理における演習環境の構築と運用

The Training Environment for IT Risk Management

川橋　裕[1,2]

Yutaka KAWAHASHI[1,2]

1 和歌山大学システム情報学センター

Center for Information Science, Wakayama University.

〒640-8510 和歌山市栄谷930

E-mail: yutaka@center.wakayama-u.ac.jp

2 大阪府立大学大学院工学研究科

Graduate School of Engineering, Osaka Prefecture University

〒599-8531 大阪府堺市中区学園町1番1号

ネットワークおよびサーバ系システムを基幹とした情報基盤において，現在はコンテンツを含めた情報基盤の整備と，社会的対応を考慮した運用管理が求められている．技術進化による多様なコンテンツは，これまでの情報基盤と一体化して運用管理の負荷を上げている．一方で情報漏洩や苦情対応など法律的，道義的責任を問われる案件が急増している．しかし，幅広い技能と経験に依存している情報危機管理において，人材が著しく不足している状況は変わっていない．これは情報危機管理における，人材育成に必要な体系化したプログラム構築が困難であることに起因する．

我々は本研究で，実践形式による演習を通じて情報危機管理の人材育成を進めてきた．実際の運用管理を模した環境下で適切な演習運用フレームを適用することが，得難い経験と座学の有効な活用につながる．本論文では，情報危機管理における演習環境の構築と運用について述べる．

It should be added content management and social interaction skill to former network and server system management. Because the content has become more complex, then the lack of social interaction, including legal action and customer service, is increasing IT Risk. However, human resources are insufficient to IT Risk management since there is not comprehensive training program.

In our research, we have been promoting the human resources development through the real scale simulations for IT risk management. This simulation is carried out under the environment constituted on the actual network, server system, complex content and social interaction. Therefore, we think that this simulation, applied the appropriate operation frame, is suitable for comprehensive training program. Then those who participate in this training will get a variety of skills and experience. In this paper, we describe the design, implementation and its operation frame of the IT Risk management training environment.

キーワード: 情報危機管理，演習環境，運用フレーム，社会的対応

IT Risk management, Training Environment, Operation Frame, Social Interaction

1 背景

情報技術の発達により，情報の発信および取得に係る情報基盤が拡大し，現在も急速な技術進化を続けている．加えて，情報基盤上で提供されるサービスの充実などにより，ビジネスや家庭でのサービス利用が顕著となっている．昨今はサービスにおける情報内容（以下，コンテンツと呼称）の多様化と複雑化，および個人情報や取引情報などへの高い機密性が求められている[1]．

従来から，情報基盤とは一般に通信基盤を指しており，ネットワークおよびサーバ系システム（以下，ネットワーク・システム）において，設備設計と運用管理設計がなされてきた．以下に，従来のネットワーク・システム運用管理に必要な要件を示す[2]

・構成管理
・性能管理
・セキュリティ管理
・アカウント（課金）管理
・障害管理

上記の運用管理要件において，構成・性能・セキュリティ管理については，技術進化によるコスト削減や高機能・高性能化が進んでいる．仮想化によるネットワーク・システム構成の変化，大容量化と高速化による性能の向上，通信パケットの検閲によるセキュリティの向上などが例として挙げられる．アカウント管理は，世界的なブロードバンド化にともなう低価格サービスの追求と管理負担の軽減が図られる一方で，障害発生などによる社会的リスクは高まっている．

従来の運用管理要件における大きな問題点は，障害管理を起点とした各要件の連携が，管理体制と人材に大きく依存していることである．現在も管理者としての人材は著しく不足しており，各要件は個別に存在することなく常に相互関係を保持している．仮想化による複雑なネットワーク・システム構成は，状況把握の困難と障害対応の遅延を起こす可能性がある．セキュリティを高めるために厳密な検閲を実施すれば，テレビ会議など大容量コンテンツ利用時に性能を低下させる．コストを意識して安価な機器を多数用いてネットワーク・システムを構成すると，障害の切り分け箇所が激増する．これらの事象はトレードオフであり，同事象の回避には，例外なく管理者の技能と経験が求められる．

さらに，現在では新たな運用管理要件が二つ求められている．一つは前述したコンテンツの管理である．Webコンテンツを構成する各種ファイルを統合化するCMS（Content Management System），XMLやHTML5による複雑なコンテンツ，および同コンテンツのサーバ分散とクラウド化などによって，ネットワーク・システムと併せた新たな情報基盤を形成している．

二つめの新たな運用管理要件は，社会的な対応の管理である．例として顧客や一般ユーザへの対応，組織内におけるセキュリティポリシの遵守と検討，および総合的な分析と提案能力などが含まれる．従来は，ファイアウォールなどでセキュリティ管理を補うなど事前防御を主体として考えられてきた．しかし，ユーザからのクレーム対応の不誠実を問われた風評被害が増えている．さらに，道義的責任の軽減など事後の対応をスムーズにするセキュリティポリシの策定や改定を迫られる．このような情報分野における危機管理を実施する体制やフレームの確立と，幅広い人材育成への希求が高まっている[3]．

本論文で述べる情報危機管理とは，コンテンツを含む情報基盤を対象としており，障害管理を起点として新たな二つを含む各要件を満たすべく対応することを指す．

上記の運用管理要件は，現在の情報基盤で運用管理に携わる人材にとって必須である．しかし，学術機関での演習科目を通じてこれらを

学習・体得することは困難である. 前述のように, 運用管理者には技量と経験が求められており, 業界で経験を積むには膨大な時間を要する[4]. 一方で学術機関には座学を活かせる十分な演習科目がない. これは演習として運用管理を総合的に実践する環境がないことと, 運用管理を学術的に体系化することが非常に困難であることに起因する.

我々は, 本研究において情報危機管理を実践する演習環境を提案する. 演習環境には設備設計と同環境の運用設計が含まれている.

本論文では, 情報危機管理の演習環境について述べるとともに, これまで実施してきた評価および今後の展望について述べる.

2 情報危機管理演習

すでに我々は, 2006年より当該環境を用いた情報危機管理コンテストを「サイバー犯罪に関する白浜シンポジウム」において2010年現在まで毎年開催している. さらに, 2008年よりIT危機管理演習として「IT keys-先導的ITスペシャリスト育成推進プログラム」で, 同様に毎年実施している. 2009年からは当該環境を遠隔利用するASP (Application Service Provider) として運用している[5].

情報危機管理演習に必要な要素について以下に述べる.

- 演習環境 (ネットワーク・システムとコンテンツ)
- 想定する脅威 (攻撃, 事故, 風評被害)
- シナリオ (演習の運用管理フレーム)
- 参加側(障害の切り分け, 苦情対応と提案)
- 運営側 (攻撃, 苦情, 情報統括責任者)
- 評価内容 (トラブルチケット, 進行表)

本論文では, 前3項目を各章に, 他を以下に示す.

参加側は, 情報危機管理演習の参加者を指す. 障害復旧作業, 電話やメールによる対応, およびマネージャによって構成される. 現在の演習環境では, 5-6名を1チームとして最大6チームまで同時に参加できる. 各チームには前提があり, 情報サービスを提供する企業に派遣された存在とする (図1). したがって, 状況報告やシステム停止の可否, および環境改善に係る提案を情報統括責任者に対して実施する (図2).

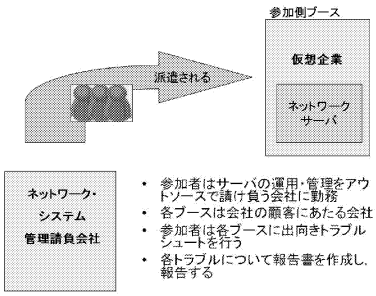


図 1 参加側の前提条件

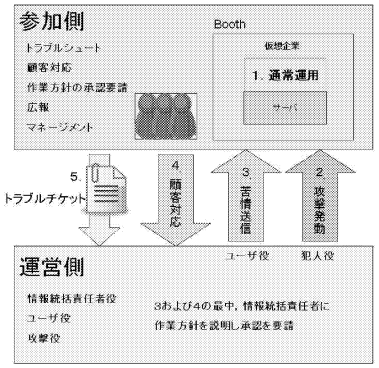


図 2 参加側と運用側の役割

図2のように, 運営側は参加側のネットワーク・システムを攻撃する. 社外の一般ユーザとして侵入して管理者権限を奪取したり, 迷惑メールを送信するスクリプトを確信犯的に設置し稼働させるなどを実施する. さらに, 当該ネットワーク・システムおよびコンテンツを利用する社内・一般ユーザとして苦

情を参加側に申し入れる．

評価内容には，障害復旧に係る作業履歴や苦情対応の内容を記載する．一般的にこれはトラブルチケットと呼ばれ[6][7]，運用管理の評価に用いられる．したがって，詳細かつ正確に状況を把握できているか，報告する相手に伝えられるかが問題となる．併せて，運営側は参加側の状況把握に努める．障害が確実に発生しているか，原因特定に迫っているか，本当に復旧したかなどを可能な限り詳細に進行表に記載し集約することで全体を把握する（図3）．

ブース1	ブース2
10:00 状況開始	10:00 状況開始
10:32 攻撃開始(Type1攻撃)	10:32 攻撃開始(Type1攻撃)
10:33 苦情電話(イマイ氏)	10:34 苦情電話(イマイ氏)
「和歌山大学のHPにとばされる」	「和歌山大学のHPにとばされる」
10:34 index.htmlをviで参照確認	（担当者名名乗る，URL確認有り）
10:42 .htaccessを同参照確認	10:39 イマイへの対応確認メール
10:47 CIOに電話連絡	10:41 .htaccessを参照
（状況説明がない，現象のみ）	10:47 CIOへ.htaccessの○○部
10:51 .htaccess全行コメントに	削除を要請，状況説明有り
10:52 担当イトウよりイマイへ連絡	10:50 自社HPに倒産と表示される
「修復したので確認して欲しい」	（○○部削除のため）
10:57 CIOより自社スポンサーサイト	10:53 イマイへ復旧確認電話
が見られないと連絡	10:56 CIOより倒産表示の苦情
（全行コメントにしたため）	

図 3 進行表の例

参加側の作業内容を把握するにあたり，演習環境では運用するサーバをVMware Serverによって仮想化している．同サーバのコンソール画面にターミナル接続して，参加側の作業内容を把握する．併せて，派遣先の企業と契約したユーザとしてサーバ内にログインし，情報収集する旨を参加側に伝えておく．

3 演習環境

本章では，情報危機管理演習に必要なネットワーク・システムについて述べる．

ネットワーク・システム構成は，参加側が担当する各ブースと運営側が担当するCIC（Contest Information Center）から成る．

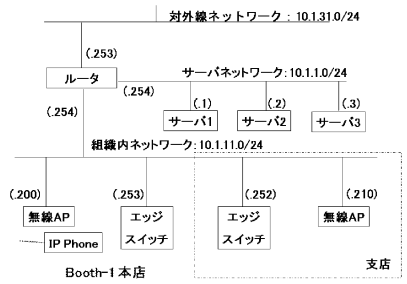


図 4 ブースの機器およびネットワーク構成

3.1 各ブース構成

各ブースには，対外線ネットワーク，サーバネットワーク，社内ネットワークが存在する（図 4）．対外線ネットワークはCICおよび他のブースと接続されている．

サーバは，DMZ（非武装地帯）となるサーバネットワークに設置される．各サーバでは社外および社内向けサービスを提供するため，ルータでアクセスを制御する．

組織内ネットワークには本店と支店が含まれている．昨今はキャリアやプロバイダによる拠点間接続サービスを用いることで，遠隔地の支店を本店と論理的に直接接続するケースが増えている．しかし，運用管理は本店側で集約されていることも多い．したがって，支店側に出向いて障害の切り分け作業をするには，運営側の情報統括責任者に許可を取る必要がある．遠隔から状況を判断し支店側で作業する必要性を納得させられなければ，支店に出向く出張費が保証されない状況を演出する．

サーバには社内用 Web サーバ，契約ユーザ用の顧客データベース，メールおよびDNS サーバが含まれている．2010 年からは社内向け認証ネットワークを本店と支店に構築し，認証に必要な Radius サーバを導入している．これらのサーバは VMware Server

によって仮想化を施している．さらに，社内利用の無線 AP（Access Point）を本店と支店に設置する．

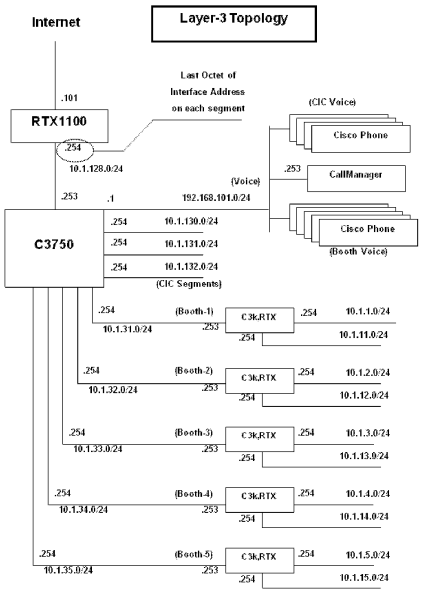


図 5 演習環境のレイヤ 3 ネットワーク構成

3.1 全体のネットワーク構成

前節のブース環境を前提として，各ブースと CIC およびインターネットを接続するツリー型のレイヤ 3 ネットワークを構成する（図 5）．演習では，本店と支店を別の部屋に割り振る一方で，各ブースのサーバを仮想化して物理的に集約する必要がある．したがって，複雑な物理ネットワークを設計しなければならない．物理的なネットワーク構成に近いレイヤ 2 ネットワーク構成を図 6 に示す．図 6 では，CIC と基幹ネットワークおよび 1 ブース間のみ示している．全体の基幹ネットワークでは，複数の論理ネットワーク（以下，VLAN）を 1 本のケーブルに集約するため，IEEE802.1q によるタグ VLAN を用いて構成している．図 6 で 2 重リングを重ねた部分がタグ接続である．基幹ネットワークには，後述する障害を発生させるために，以下のような配慮が必要である．

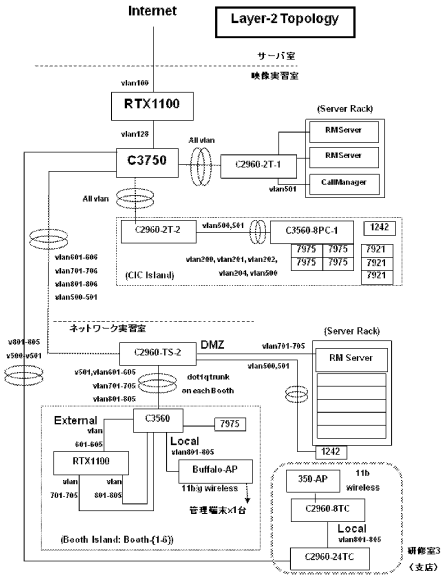


図 6 演習環境のレイヤ 2 ネットワーク構成

- ループ接続によってネットワークを混乱させる場合，基幹ネットワーク全体を使用不能としないために，各 VLAN 内のブロードキャスト流入出量を制限する
- 参加側と運営側を結ぶ IP 電話を設置し，ブース内が通信不能でも通話可能とする
- 運営側の攻撃を防ぐためルータ内のアクセス制御を施しても，攻撃元を変えて継続できるよう CIC に複数のネットワークを用意する
- 経路情報の攪乱で基幹ネットワーク全体を混乱させないため，基幹ネットワークを静的経路で運用する

これまで実施してきたコンテストや演習では，事前に参加側へ上記の情報をすべて

開示している．開示する情報には，ネットワーク・システム構成だけでなく，提供サービスの内容，OS やサーバ・プログラムの種類やバージョン番号，および管理者情報など，通常運用に必要な情報が含まれている．

4 想定する脅威と障害内容

脅威とは攻撃手段を指す．攻撃手段は仕込みとトリガ（発火手段）の 2 つの段階で構成される．本演習環境では，演習を開始する時点でネットワーク・システムおよびコンテンツは通常運用されている．しかし，環境内にはすでに攻撃手段が仕込まれているため，トリガによって障害が発生する．これらは後述するシナリオの一部である．

4.1 仕込み

攻撃手段に応じて，運営側あるいは参加側に脆弱性や設定の不備などを施す．昨今頻発している諸外国からの bot によるパスワード攻撃など，サーバ上で不正と判断し難いアクセスによってユーザや管理者の権限を奪取することは運営側の仕込みとなる．リロード攻撃でサーバを圧迫することや，苦情を出すユーザが所有するパソコンの設定不備も，運営側の仕込みとなる．

一方で，OS やサーバ・プログラムの脆弱性を用いて bot を設置し，第 3 者にステップ攻撃する事案も増加している．データベースのサニタイズ未処理などによる情報漏洩も同様である．メールサーバ設定の不備で第 3 者転送が可能な場合，気づかないうちに迷惑メール転送サーバとなってしまう．これらの脆弱性や設定の不備は参加側にあらかじめ施しておく仕込みである．

4.2 トリガ

トリガとは，仕込みを用いて障害を引き起こす行為を指す．例として，仕込んだ脆弱性を用いて管理者権限を奪取し，ホームページや設定を改ざんする．情報漏洩では，サニタイズ未処理を悪用する SQL 構文の入力で奪取した情報を，演習環境内の掲示板に掲載する．これらは複数の段階を経て発火させるトリガである．

リロード攻撃や SYN Flood 攻撃などの物量攻撃，および支社となる部屋内でケーブルをループ接続させる場合は，行為自体が単体のトリガとなる．

仕込みとトリガは常に 1 対 1 の関係ではない．改ざんが管理者権限の奪取から連なる行為で発生したのであれば，改ざんされたホームページや設定を書き戻しても根本解決にならない．したがって，再度の発火が可能であり，他の改ざんを発火できる．

4.3 障害内容

ホームページが見えない，つながらない，メールが送信できないといった障害内容は表面的な事象であり，多くの苦情には抽象的な情報しか含まれていない．苦情を寄せるユーザに対して的確かつわかりやすい質問で対応し，具体的な情報を取得する必要がある．復旧作業と苦情対応が連動しなければ，効率よく復旧できない．一例として，実際に実施した一連の流れを以下に示す．

- ホームページが見えない or 表示が遅い
- Web サーバへのアクセスが多く高負荷
- アクセス元が多く制限が困難
- 掲示板への大量書き込み「祭り」が発生
- 書込内容は顧客データ（自社の情報漏洩）
- ログから不正な SQL 構文の入力を確認
- サニタイズ未処理の部分が存在する

上記のように最短でたどり着ければ良いが，実際に各項の間には多くの可能性があり，適切な技能と膨大な調査が求められる．

4.4 その他

「必ず発火させなければならない」「必ず復旧できる障害内容でなければならない」「仕込みから障害内容につながる流れのパターンをいくつか用意する」．これらを実現することは非常に困難である．しかし，復旧できない障害を発生させては意味がないため，我々はログファイルを改ざんしない．必ず痕跡を残しておいて，この痕跡に気づかせるよう留意する必要がある．

一方で，OS や各サーバ・プログラムが出力するログファイルに，攻撃に関するログのみが残っている状態は避けるべきである．我々は，Web サーバのコンテンツ巡回プログラムを作成した．本プログラムは，設定した頻度で閲覧や投稿を自動的に実施する．さらに，SSH（Secure SHell）などのターミナル接続でも，自動的にアクセスしてサーバにログを残すクライアントを構築し，正規のサービス利用による通常運用の状態を演出している．

参加側は，実際に蓄積される膨大なログの中から不自然なアクセスを見つける必要に迫られる．

上記のように，想定する脅威と障害内容には複雑な配慮が必要である．同様に，これらは次章で述べるシナリオの一部となり，障害復旧および演習の評価につながる．

5 シナリオ

前章で述べた仕込み，トリガ，および障害内容に加えて，シナリオ要件には対応予測と評価指標がある（図 7）．対応予測と評価指標にも，前章で述べたように配慮が必要である．運営側から苦情を申し出た段階では，提示する情報を少なくする必要がある．参加側が詳細かつわかりやすく質問しない限り，障害内容に関する詳細情報を提示しない．しかし，状況が膠着することは好ましくないため，適度な時間をとって徐々に詳細情報を提示する．

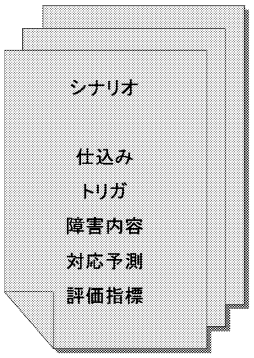


図 7 シナリオ要件

5.1 対応予測

運営側の CIC では，参加側の各ブースに対して 1-2 名のブース担当者を決める．各担当者は，2 章で述べたように，担当するブースへの攻撃や苦情，および情報統括責任者としての役割を担う．しかし，担当者の個人の技能と経験に差異が生じると，すべての参加側に対して適切な評価が出来ない．特に障害内容の伝達や，情報統括責任者としての判断などには，各担当者の対応内容に均質化が求められる．

したがって，事前にシナリオを作成し，障害復旧に至るまでの流れを運営側全体で共有しておく必要がある．すべてのシナリオにおいて，上記の流れをフローチャートで作成し，各自の手順を確認する（図 8）．

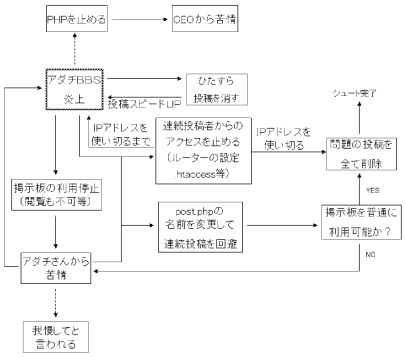


図 8 シナリオのフローチャート

図 8 に示した事例は，掲示板に連続して書き込む迷惑行為シナリオの簡易フローチャートである．連続書き込みは，リロード攻撃などと同じく，TCP/IP 通信上は正規のアクセスとなるため，あらかじめ状況を把握することが難しい．しかし，ユーザすべてに対して連続書き込みを一方的に止めることが免責事項となってない場合，情報統括責任者の許可が必要となる．

実際のフローチャートはさらに複雑である．複数の仕込みやトリガを要するシナリオでは，参加側の対応パターンに応じた条件分岐を用意する．

発生する障害への対応予測には，障害復旧の手法および障害の切り分け箇所が含まれている．参加側の対応パターンと，各パターンに対する運営側の対応をあらかじめ予測して，これを実施，確認することは，運営側にも人材育成の機会となる．

5.2 評価指標

基本的に障害復旧の達成度と時間が演習の評価対象となる．さらに，詳細かつわかりやすいトラブルチケットの記述内容が含まれる（図 9）．運営側から見た，苦情への対応と詳細情報の提示の度合いも加味される．運営側からの提示の度合いが低ければ評価が高くなる．

加えて，トラブルチケットへの補足事項も対象となる．すなわち，障害復旧の後，今後の対応として利用規程や免責事項の改定，サーバ側のアップグレードなど，リスクを説明した上での改善提案があれば，内容を精査して評価に加える．

--Booth2
1.障害の概要 ID:01
発生時刻 :10:36
発生報告 発見者 :イマイ 様
担当者 :カスタマー担当 中村
内容 :
ホームページへアクセスしたところ、異なるページへ転送されてしまった。転送先は、和歌山大学であった。

-- 2.障害への対応 [2009-09-05]
[10:34]イマイ様より上記内容の障害報告をメールで受理。
[10:36]イマイ様より上記内容の障害報告を電話で受理。
[10:38]障害の確認。
[10:39]イマイ様宛へ原因究明の旨のメールを送信。
[10:41]apache内の◆◆に、和歌山大学 (http://www.wakayama-u.ac.jp)が指定されているのを確認

[10:50]CEOへ障害の発生を連絡。
[10:51]対象箇所の削除を実行し、問題を解決した。
[10:52]イマイ様へ対応に関するご連絡の電話。
[10:53]イマイ様へ問題解決の報告メールを送信。

-- 3.原因の究明
設定ファイルを確認した結果、和歌山大学のページへ転送する設定になっていた為、
CEOの同意を得て該当箇所の削除を行い問題の解決をした。

図 9 トラブルチケット例

6 評価

本章では，これまで提示した情報危機管理における演習環境の構築と運用について評価する．

人材育成の観点からは，技能や経験の積み重ねを定量的に評価することは困難である．本環境は，大学の講義などのように，参加者に対して一定期間定常的に演習を繰り返していない．

一方で，一般に公開されているいくつかの演習形式と比較して，定性的に評価することは可能である．Black Hat が主催する Black Hat Training や，日経バイトが主催したセキュリティ・スタジアムが類似した演習形式である．しかし，技術偏重であることや，攻撃する側の競技が含まれることなどの違いがある．本演習環境は，上記と

比較して特に以下の点で優れている.

- シナリオに沿ったロールプレイ形式
- 通常状態から実際に攻撃する
- 社会的な対応を含めたシナリオと評価
- 上記を実現する環境と運用フレーム

さらに，情報危機管理コンテストを併催している「サイバー犯罪に関する白浜シンポジウム」では2008年に経済産業大臣賞を受賞しており，受賞理由に同コンテスト名が明記されている[8]. IT 危機管理演習を実施している「IT Keys-先導的 IT スペシャリスト育成推進プログラム」の平成22年度中間評価結果では，「世界最高水準の IT 人材育成に向けた成果」との評価を受けた. 本演習は「実際におきうるインシデントとその事後処理までの情報システム管理者としての実践に即したロールプレイ形式の演習科目」として認められている[9].

なお，IT危機管理演習では，事前の座学に加えて，演習後に派遣先企業の役員向けを想定したプレゼンテーションを実施する. 参加側は，派遣 SE として発生した障害と対応内容を説明し，今後の対策と必要な契約金額の上乗せを提示する. 運営側は，本プレゼンテーションも演習の結果として評価する.

7 今後の課題

本演習環境において最大の問題点は，実践に即した演習環境の提供が可能になった反面，すべての機材と人員が会場に赴くことである. 一度に30人程度が参加できるとはいえ，機材を1カ所に集める仕様にはデメリットがある. これまでに述べたコンテストや演習において搬入出する機材の総重量はおよそ900kgであり，定常的に運用するには遠隔参加が必須である.

我々はすでに ASP 化に向けた設計と運用を開始しており，2009年より情報危機管理コンテスト予選において遠隔環境を実施している[5]. 予選では PPTP(Point-to-Point Tunneling Protocol) を用いて参加側の端末を遠隔参加させる. しかし，ネットワーク障害などのシナリオを適用すると演習自体が成り立たなくなるなどの問題がある. 現在はインターネット上で L2TP (Layer 2 Tunneling Protocol) など下位レイヤのトンネル接続を実施することで当該問題の回避を検討している. 併せて，専用回線における IEEE802.1q の多重カプセル化にも着手している. 遠隔参加による演習を維持するための運用フレームについても，さらに検討する必要がある.

次に進行表の問題がある. 運営側に十分な人員がある場合を除き，参加側の各チームに1-2名が担当すれば，一人あたりの役割が増大する. 特に現在は演習の実施中に進行表を担当者が作成するので，さらに負荷が高まる. 我々は，進行表の作成支援として，現在 UNIX サーバシェルのヒストリを逐次監視する機構を作成して負荷軽減を図っている. 参加側の行為を実行コマンドで把握し，これを進行表に埋め込む. 今後は管理対象の仮想サーバ内に chroot や jail でゲスト空間を構築し，ホスト空間から構成や設定の変更を監視できるシステムの設計に取り組む.

さらに，障害および障害対応に係る損害評価を詳細に実施することで，より実践的な PBL (Problem Based Learning) や PML (Problem Management Learning) とする.

8 まとめ

本論文では，「事故前提社会システム」

に対応できる人材育成という方針[1]に沿った情報危機管理の演習環境の構築と運用について述べた．コンテンツを含む情報基盤では，運用管理に多様な技能と経験が必要であるため，社会的対応を含めた要素を演習環境に取り入れている．さらに，人材育成の観点から，詳細なシナリオを策定し常に参加側との対話によって原因にたどり着かせるというユニークな運用フレームを実施している．すでにシナリオ数は 20 を超え，大学など学術機関における半期の演習に応えられるだけの蓄積がある．

筆者を含め，筆者の研究室は運用管理研究に携わるとともに，現実の運用管理にも従事している．したがって詳細なシナリオを策定し得るとともに，参加側の対応予測が可能となっている．さらに，セキュリティ事案でない場合でも重大な被害を引き起こす可能性を十分に承知している．

今後は，限定された環境下での人材育成にとどまらず，汎用的かつ実践的な演習環境の構築と運用によって，さらに広く人材を育成できるよう努める所存である．

参考文献

[1] 経済産業省 ：「情報セキュリティ総合戦略」，
http://www.meti.go.jp/policy/netsecurity/downloadfiles/Strategy_Summary.pdf
[2] Allan Leinwand, Karen Fang：「Network Management: A Practical Perspective (2nd Edition)」, Addison-Wesley, 1995
[3] 中野宏幸，大林厚臣：「IT 障害に関する分野横断的演習の取組み：分野を超えた情報共有と連携協力の仕組みづくりに向けて」，社会技術研究論文集，5，pp. 143-155，2008.
[4] 南条 優：ネットワーク運用・管理の重要性と人材育成（マルチメディア時代のネットワークエンジニア育成法<特集>）．ビジネスコミュニケーション 26(7)，pp. 26-31，1989.
[5] 塩崎康平，川橋 裕 ：「情報セキュリティ技術者・管理者育成を目的とした情報危機管理演習の環境構築と運用支援」，学術情報処理研究論文集 ，No. 14 ，pp. 117-128 ，2010.
[6]泉 裕，齋藤彰一，上原哲太郎，國枝義敏 ：「柔軟なネットワーク管理フレームワークを提供するトラブルチケットシステムの構築」，<特集>インターネットアーキテクチャ技術論文，電子情報通信学会論文誌．B，通信 J86-B(8)，pp. 1502-1514，2003.
[7]亀谷信行，泉 裕，上原哲太郎，國枝義敏 ：「トラブル対応支援システムの構築」，情報処理学会研究報告．DSM，[分散システム/インターネット運用技術] 2000(113)，pp. 61-66，2000.
[8] 経済産業省，(2008)，平成 20 年度情報化促進貢献個人等の表彰
http://www.meti.go.jp/press/20080917001/20080917001-2.pdf
[9] 文部科学省，(2010)，先導的ＩＴスペシャリスト育成推進プログラム（平成 19 年度採択）中間評価結果
http://www.mext.go.jp/b_menu/houdou/22/03/__icsFiles/afieldfile/2010/03/30/1292179_1.pdf

論文

タスク種別とユーザ特性の違いが Web 情報探索行動に与える影響: 眼球運動データおよび閲覧行動ログを用いた分析

Influence of Task Type and User Group on Web-based Information Seeking Behavior: Analysis of Eye Movement Data and Client-side Browsing Logs

高久雅生[1*], 江草由佳[2], 寺井仁[3], 齋藤ひとみ[4],
三輪眞木子[5], 神門典子[6]
Masao TAKAKU[1*], Yuka EGUSA[2], Hitoshi TERAI[3], Hitomi SAITO[4],
Makiko MIWA[5], Noriko KANDO[6]

1 物質・材料研究機構 科学情報室
National Institute for Materials Science
E-mail: TAKAKU.Masao@nims.go.jp
2 国立教育政策研究所 教育研究情報センター
National Institute for Educational Policy Research
E-mail: yuka@nier.go.jp
3 東京電機大学 情報環境学部 情報環境学科
Tokyo Denki University
E-mail: terai@sie.dendai.ac.jp
4 愛知教育大学 教育学部
Aichi University of Education
E-mail: hsaito@auecc.aichi-edu.ac.jp
5 放送大学 ICT 活用・遠隔教育センター
The Open University of Japan
E-mail: miwamaki@ouj.ac.jp
6 国立情報学研究所 情報社会相関研究系 / 総合研究大学院大学
National Institute of Informatics / Graduate University for Advanced Studies
E-mail: kando@nii.ac.jp
*連絡先著者 Corresponding Author

Web 利用者の情報探索行動の理解のために，眼球運動データ，ブラウザログ等の情報を包括的に用いて，ユーザ特性，タスク種別の違いがいかに探索行動に影響しうるかを検討した．ユーザ特性として図書館情報学専攻の大学院生と一般学部生の 2 グループを設定し，それぞれ大学院生 5 名，学部生 11 名が実験に参加した．タスク種別としてレポートタスクと旅行タスクの 2 つを設定して，それぞれ 15 分間ずつの Web 探索行動を観察した．実験の結果，タスク種別の影響として，サーチエンジンの検索結果一覧ページ上における視線注視箇所として，旅行タスクではスポンサーリンク，レポートタスクではスニペット領域がより多く見られるなど，タスク毎に着目する情報が異なることが示された．また，旅行タスクではサーチエンジン検索結果一覧ページから 2 回以上たどったページの閲覧回数がより多く，異なるタスクにおいて情報収集方略が異なる傾向が示唆された．一方でユーザ特性の影響として，大学院生は学部生に比べて，ページ閲覧をすばやく行い，タブ機能を用いた並列的な閲覧行動も観察されるなど，効率的な情報収集を目指した行動の差異が見られた．

We investigated the effect of user-based and task-based properties of people's information seeking behaviors by using screen-capture videos, browser logs, and eye movement data. The participants were 5 graduates and 11 undergraduate students. They were given two tasks: a "report task" in which they had to gather information to prepare a report on world history, and a "trip task," in which they had to plan for a trip. The results of the analysis of eye movement data show that there are several differences in focused areas for each task; e.g., participants viewed snippet blocks on the search engine results pages in the report task more often than those in the trip task, and they viewed sponsor link blocks more often in the trip task than those in the report task. Furthermore, the results of both participant groups showed that for the the trip task, participants tended to search deeper for pages through numerous links. On the other hand, graduate students tended to have more parallel browsing behaviors, i.e., using tab browsing functions, compared with the undergraduate students.

キーワード: Web 情報探索行動，Web 行動カテゴリ，眼球運動データ，Link depth
Keywords: Web information seeking behavior, Web action category, Eye movement data, Link depth

1 はじめに

Web の爆発的な普及とともに，Google に代表されるサーチエンジンが日常的に使われるようになり，検索エンジンの性能向上と探索行動の支援が，より重要となってきている．多様なユーザの要求にこたえうるサーチエンジンが求められている [1].

たとえば Marchionini は，現在のサーチエンジンが満足している情報要求と満たされていない情報要求とを対比して，調査学習型探索（exploratory search）の重要性を指摘している [1]．調査学習型探索とは，事実検索や質問応答のように 1 回の質問で答えが得られる簡単な検索ではなく，探索のゴールを少しずつ明確化しながら新しい知識を獲得していく学習や調査における探索である．

これまで，一般の利用者が Web 全般もしくはサーチエンジンをどのように利用しているかについては，ログ分析やユーザ実験，インタビュー調査などを通じた研究が行われている [2]．しかしながら，情報探索の過程で情報をいかにピックアップし，その結果として次にどのような行動が選択されるのかについて，ユーザの情報探索行動全体にかかわる要因との関係について，包括的に検討している研究は数少ない．

そこでわれわれは，Web 利用者の調査学習型の情報探索行動を理解するために，ブラウザログ，画面キャプチャー映像，発話，眼球運動，インタビューなどの様々なデータを収集し，探索タスクや経験の違いによる探索行動への影響を検討する．収集データをもとに，ユーザが閲覧したページ種別毎の閲覧回数や閲覧時間，サーチエンジンにおける眼球運動に対する分析やクリックした文書ランクの分析を行い，タスクや経験による探索行動の特徴を明らかにしてきた [3][4][5][6][7][8]．また，インタビューやプロトコル発話の質的分析によって，どのように利用者が探索を進め，新たな知識を獲得，既存知識の訂正，限定，関連付け，詳細化，概念を異なる枠組みでとらえなおす転換などが起こっているか明らかにしてきた [9][10][11][12]．本稿では，閲覧行動データの分析に加え，眼球運動データの分析を通じて情報探索行動の精緻な理解を図ることを目指す．

以下，2 章では関連研究を中心に背景と目的を述べ，3 章では実験方法と手続きを示す．4 章で実験から得られたデータの分析手法を説明したあと，5 章においてその分析結果を示す．6 章で考察と議論を行い，最後にまとめを述べる．

2 背景と目的

従来の Web 情報探索行動の研究においても，多くの研究手法が用いられてきた．それらは主

として定性的分析と定量的分析の 2 者に分けられる．定性的，定量的いずれの研究手法にあっても，タスク種別やユーザ特性をとりあげた研究がいくつか見られる．以下，2.1 節では，タスク種別の影響を調査した研究の動向について述べ，2.2 節ではユーザ特性との関連を調査した従来の研究について述べる．また，2.3 節において眼球運動情報を用いた Web 情報探索行動の研究について述べる．

2.1 タスク種別

Kellar ら [13] は Web 情報探索を以下の 5 つの種別に分類している: 事実発見（fact-finding），情報収集（information gathering），ブラウジング（browsing），トランザクション（transaction），その他（others）．そのうえで実地の状況に近い設定でのフィールド調査を行い，「その他」に分類される種類のタスクを除くと，全探索活動における各タスクは 13.4%〜46.7% の比率で行われていることを報告したうえで，タスク種別のうち，情報収集・ブラウジング間と，ブラウジング・事実発見間では遂行時間に有意な差があったと述べ，タスク種別毎の質的な差異が Web 情報探索におけるタスク遂行時間に影響を与えていると結論付けた．

タスクの特定性および複雑性も探索の遂行に影響を及ぼす要因となる [14]．Kim ら [15] はタスクの違いと利用者の認知特性の違い，Web 検索性能の違いについて研究を行っている．Kim らはタスク種別を主題検索タスクと既知事項検索タスクの 2 つに分類したうえで実験を行い，タスク種別が再現率，適合率の双方に有意な影響を与えたと報告した．既知事項検索タスクでの検索性能は，再現率，適合率ともに主題検索タスクにおけるものよりも高かった一方で，主題検索タスクにおける，探索時間，閲覧ページ数，リンクをたどった回数がいずれも既知事項検索でのそれを上回ったと報告している．

Thatcher[16] は，タスク種別と，Web 上での探索経験にともなう認知的探索戦略との関係を調査した．Thatcher は事実発見（directed fact-finding search）とブラウジング（general purpose browsing）の 2 種類のタスク [17] を設定した実験を行った結果，事実発見タスクでのみ，Web 経験のもっとも豊富な実験参加者がその他の参加者よりも統計的に有意に多くのクエリを発行するという違いが観察されたと報告している．

一方，タスク概念は，情報要求，入手過程，探索結果という個々のプロセスごとの複雑さに従っても分類できる．Byström ら [18] は，質問紙調査および構造化インタビューに基づく定量的な調査法により，タスクをその複雑さに応じた 5 つのカテゴリに分類している．

以上に挙げたように，従来の研究は多様な目的と分析手法によるものがあるが，その焦点は，Web 探索タスクと，そのタスクに基づく情報要求と，さらにはそれらの情報探索行動に対する影響といった点に置かれており，タスクの特徴とそのタスクが探索行動に与える影響とを解明することが，近年の Web 研究において重要な研究トピックの一つとなっている．とくに，今日の Web サーチエンジンに代表される検索行動での情報要求は，主題検索を扱うにとどまらないものとなってきている．たとえば，Broder[19] はサーチエンジンのクエリログとユーザアンケートを通じて，Web 上での情報探索タスクを情報指向（informational），ナビゲーション指向（navigational），トランザクション指向（transactional）の 3 種類に分類し，サーチエンジン上でのそれぞれのタスク比率を報告している．しかしながら，このような個々のタスクの特徴が実際のユーザの行動にどのような影響を与えているかは必ずしも明らかとなっていない．

また，既存の研究では，主題検索タスク（または情報指向タスク）と既知事項検索タスク（またはナビゲーション指向タスク）との比較や，固有の検索リソースやシステムを想定してその中での探索行動に限定した状況での比較といったものが多い．特定の検索システムに対するキーワード検索結果を数件ながめて求める文書にたどりつくという程度にとどまらず，調査学習型探索で必要となる，複数の検索リソースやシステムの選択と比較をおこない，かつ，検索結果から得られた情報をさらに詳細化して何度も検索を試行しながら知識獲得と訂正がおこなわれるような，より広範なタスク遂行の範囲を視野に入れた研究は少ない．本研究では，Web 上で日常的に行われる，典型的な情報収集タスクとして 2 つのタスクを設定し，Web 上での対話的な情報探索タスク全般に関わる要因を探るため

の実験を行う．

2.2 ユーザ特性

ユーザの特性が情報探索行動にどのような影響を与えているかは，ユーザの効果的な学習を支援するうえでも，検索システムの改良や対象ユーザの設定のうえでも重要なテーマのひとつであり，多くの研究がなされてきた．この際，ユーザ特性を取り上げた研究は主に，1）探索する対象に関する主題知識，2）検索環境に対する経験，3）問題解決スタイル，の 3 つに大別される．

主題知識レベルの異なる利用者の探索行動を取り上げた研究の一例に，Hembrooke ら [20] によるものがあり，主題知識とシステムに対するフィードバックの関係を調べるためユーザのタスク遂行の度合および使用方略を分析し，主題知識の差が使用クエリの詳細度や検索方略の違いとして現れることを報告している．Moore ら [21] は，ユーザの経験を取り上げた既往研究をレビューしたうえで，経験・知識・専門性としてとらえられてきたカテゴリを，1）検索経験，2）オンラインデータベース検索経験，3）情報探索経験，4）Web/インターネット経験，5）コンピュータ経験，6）その他，の 6 つに大別できるとしている．

Holscher ら [22] は，Web 経験を持つ被験者を対象とした情報探索実験と，経済分野の主題知識を持つ被験者を対象とした実験とを組み合わせて分析し，双方の経験・知識を持つグループが探索タスクの成功に結び付くといった傾向だけでなく，Web 経験が高く主題知識の低いグループのユーザはサーチエンジンへのクエリ形式で詳細な検索指定を行う傾向があるなど，それぞれの経験・知識を持つグループごとに異なる方略を使っている可能性を指摘している．また Lazonder ら [23] は，高校生のうち Web 探索経験のある学生グループとほとんど無い学生グループとを対象とした探索実験を通じて，探索経験の多いグループは他のグループと比較してサイトの所在を探すタスクにおいて有効な探索ができた一方，サイト内の情報を探すタスクにおいては両者の違いがなく，探索方略に対する経験の差による影響の違いを示唆している．

そこで本研究では，Web 上での情報探索タスクに必要な 4 つの知識: 1）検索システムの機能および使い方，2）Web 上の種々の情報資源の機能および使い方，3）各タスクの探索方略，4）各タスクの個別具体的トピックに関する主題知識，のそれぞれに対する知識およびその活用経験に差があると考えられる，図書館情報学専攻の大学院生と一般学部生という 2 種類のユーザグループを設定して実験を行うこととした．

2.3 眼球運動

眼球運動データは，ひとがどの部分を見ているかまでを捉えられる，もっとも精緻なデータの一つとして，幅広い分野で研究されてきた．たとえば，認知科学における問題解決研究や，ヒューマンコンピュータインタラクションの研究分野においては，眼球運動データを用いた分析手法を使って，環境から提供される外的な資源のうち，人がどのような情報に注目し，またそれらの情報を課題の解決にどのように利用しているのかについての研究 [24][25] がある．このような眼球運動データは，ユーザが注目する視点をとらえる生理的側面からのフィードバックであり，たとえば，学習者の眼球運動データを用いて学習者の行動を理解し，学習者に適切なフィードバックを与えるシステムを開発する試みも報告されている [26]．

一方，情報検索研究では，検索システムが提供する情報オブジェクトのどの部分に着目しているかを把握し，探索過程における提供情報の影響をはかることが重要となる．たとえば Kelly[27] は，検索システムに対する適合フィードバックという観点から，ユーザの検索システムに対する自然なインタラクションを観察することによって得られる情報を使った擬似適合フィードバックの重要性を指摘し，眼球運動データの活用によりユーザに有益な情報を提供できると提案している．近年では，眼球運動データは Web ユーザビリティ調査などでの利用も一般的なものとなっており [28]，情報検索研究においてもとりわけサーチエンジンの検索結果一覧ページ (search engine results pages; 以下，SERP と呼ぶ）に着目した研究が多くなされるようになってきている [29]．

Granka ら [30] は眼球運動データを使って SERP におけるユーザのふるまいを調査している．Granka らは，5 つのナビゲーション指向タスクと 5 つの情報指向タスクとを用い，実験参加者の Google サーチエンジンでの 3 分間の検

索タスク遂行の様子を観察した．眼球運動データをもとにして実験参加者の SERP 上の検索ページの情報をどのように見て，複数の SERP をどのように使ったかを分析した．分析結果から，実験参加者は SERP 上の 1 位および 2 位にランクされたページの情報を同程度ながめ，上位から下位のランクに線形に読んでいったとの結果を報告している．

Guan ら [31] はユーザの検索行動が正解ページ情報の表示位置によって影響されるかを調査した．実験では情報指向タスクとナビゲーション指向タスクの 2 種類を用い，SERP 上の表示位置をコントロールした．実験参加者のパフォーマンスの分析結果から，とくに情報指向タスクにおいて正解ページの表示位置を下げると，ユーザのタスク遂行時間はより長くなり，正解ページの発見率も下がると述べ，くわえて視線停留データの分析結果からは，正解ページの表示位置が上位にある場合，正解ページの注視率と注視時間がともに上がることを報告している．Pan ら [32] も，実験参加者の適合性の判断に Google のランキングがどれだけ影響を与えているかの実験を行っている．

さらに，Lorigo ら [33] は眼球運動データを使って検索行動におけるタスクと性差の影響を調べた．この実験では，5 つのナビゲーション指向タスクと 5 つの情報指向タスクとを用い，実験参加者の Google サーチエンジンでの 2 分間の検索タスク遂行の様子を観察した．実験の結果からは，タスク種別がページ閲覧時間の長さおよび視線停留回数に影響を与え，性差が SERP のスキャン方略に影響しているとの分析が示された．

最後に Lorigo ら [29] は，Google サーチエンジンにおける眼球運動データと探索行動の比較を行ってきた研究プロジェクトをまとめたうえで，Yahoo!サーチエンジンも用いてサーチエンジンの違いによる分析結果の違いを確認する結果を報告した．その分析結果によれば，Google と Yahoo!の間では眼球運動データの結果に違いはほとんど見られなかったと報告され，先行する探索行動の分析結果は Yahoo!に対しても適用可能ではないかとの示唆を与えている．

以上の眼球運動データを使った研究に共通する特徴には，1）検索タスク種別としてナビゲーション指向タスクと情報指向タスクの 2 種類を用いている点，2）分析対象として SERP ページにおけるユーザの行動に焦点をあてて分析している点，の 2 点が挙げられる．しかしながら，Lorigo ら [29] が既に述べたとおり，この眼球運動データを用いた情報探索行動の研究においてはさらなる知見の蓄積が必要とされている．なぜなら，眼球運動データの分析手法だけでもまだ未成熟な点があり，そもそもどのような分析が可能であるか，他の研究との比較が可能であるか，といった点で合意が得られているとは言いがたい．さらに，他のユーザビリティ研究等で用いられている研究手法との統合に向けても課題は多い．本研究ではできるだけ自然な状況での探索行動を観察することによって，より豊かなユーザ行動の理解を得ることを目的とし，かつ，サーチエンジン利用に特化した状況だけでなく，検索結果の先にある Web ページでの利用者の探索行動をも視野にいれる点が上述の先行研究との違いとなる．

2.4 研究の目的

次節以降に述べる探索実験をおこなうにあたって設定した，本研究の研究目的を以下に挙げる．

- ユーザの Web 上での探索行動への経験や習熟度に応じて，探索行動に違いが観察されうるか？
- 探索タスクの種類に応じて，Web 上での探索行動には違いが観察されうるか？

3 実験デザイン

3.1 実験参加者

大学院生 5 名（男性 4 名，女性 1 名，年齢 23 歳から 28 歳），学部生 11 名（男性 5 名，女性 6 名，年齢 19 歳から 21 歳）が実験に参加した．大学院生は全て図書館情報学を専攻していた．学部生の専攻は経済学，文学，工学，外国語（スペイン語），心理学，理学，土木工学など多様であった．

実験は 2007 年 11 月（学部生 11 名），2008 年 3 月（大学院生 5 名）の 2 回にわけて実施し，実験参加者には謝金を支払った．

3.2 タスク

1 章でも述べたように，本研究は調査学習型探索に着目した研究を行っている．そこで，何度も検索を繰り返し，少しずつ知識を獲得していく様子を観察できるタスクとなるよう意図して，探索タスクを設定した．

これまでのわれわれの経験 [10][11][12] では，特定の答えを探すような事実応答や質問応答のタスクや，特定の細かなトピックをあらかじめ設定したタスクを実験参加者に与えた場合，タスク開始からほどなくして，実験参加者の何人かは行動が停止してしまい，探索行動の観察が十分に行えないことがあった．逆に，実験参加者にテーマを選んで実行してもらう場合には，設定したタスク実行時間を充足するかたちで探索行動を継続できた．これは，実験参加者自身の関心に基づくテーマであるため，実験参加者のモチベーションを高く保てること，もともと実験参加者があらかじめそのテーマであれば遂行できそうだとの判断によって設定したテーマであることが多いことなどに起因すると著者らは考えている．

このような探索の際の利用者自身の状況に合わせた動機付けの重要性について，たとえば，Saracevic は適合概念のレビュー [34][35] において，「システム適合性（System relevance）」「主題適合性（Topical relevance）」「認知適合性（Cognitive relevance）」「状況適合性（Situational relevance）」「感情的適合性（Affective relevance）」という 5 つのレベルでとらえられるとまとめたうえで，このうち動機付けは 5 つ目の感情的適合性にあたるとし，他の適合概念に対して基底的な関係を持つと述べている．

そこで本研究では，タスクにおいて対象とする具体的なトピックは実験参加者が設定できるものとした．これは，実験参加者が自身の情報要求と考えて高いモチベーションを持ちながら探索でき，ひいては，設定したタスク実行時間の間，参加者が探索行動を継続でき，それらの様子を観察，記録できることを重視したものである．

今回の実験では，レポートタスクと旅行タスクの 2 種類のタスクを設定した．これら 2 種類のタスクはどちらも必要な情報を探索し収集するタスクであった．レポートタスクは，「世界史」のレポートを書くために必要な情報を探索し収集するというタスクであった．実験参加者にとって親近性のあるタスクとするため，高等学校の必修科目である世界史を選択し，時代や地域等の具体的なテーマは実験参加者が興味のある対象を選んだ．一方，旅行タスクは，身近な人との「旅行計画」のために必要な情報を探索し収集するというタスクであった．レポートタスクと同様に，具体的な目的地や同行者等の設定は実験参加者が選んだ．また，実験参加者には探索遂行中に発話するよう求めた．以下に，レポートタスクと旅行タスクのインストラクションを示す．なお，下線部（＿＿＿＿＿＿）には実験参加者が選択した具体的なテーマを挿入してから探索を開始した．

レポートタスク:

> 大学の一般教養の授業で，世界史を対象に自分の興味のあるテーマについてのレポートを書く課題が出ました．テーマは，＿＿＿＿＿＿ にしました．それでは，レポート作成の事前調査としてインターネットを使って関連資料を集めましょう．調査に使える時間は 15 分です．役に立つサイトを探しましょう．課題の情報が掲載されているページをブックマークに追加してください．

旅行タスク:

> あなたは，＿＿＿＿＿＿ と行く旅行を計画することになりました．時期は ＿＿＿＿＿＿ で，期間は ＿＿＿＿＿＿，場所は ＿＿＿＿＿＿ です．一緒に行く人たちに教えてあげるつもりで，その地域への交通手段，行ってみたい場所や行事などについてインターネットを使って調べましょう．調査に使える時間は 15 分です．役に立つサイトを探しましょう．課題の情報が掲載されているページをブックマークに追加してください．

表 1 に，実験参加者が実験時に実際に選択したテーマおよび設定の具体例を挙げる．

3.3 手続き

図 1 は実験手続きの流れを示している．実験参加者は事前アンケートとして，Web サーチエ

表 1　実験参加者がレポートタスクにおいて選択したテーマ，および旅行タスクにおいて選択した設定の具体例（抜粋）

	レポートタスク	旅行タスク
大学院生 ($n = 5$)	アメリカの黒人, 仏教成立, ユーゴスラヴィア解体	大学の友人 4 人と春休み頃に 2 泊 3 日で東北旅行, 友人 2〜3 人と 3 月に 3 泊 4 日で屋久島旅行, 彼女と二人で夏に 2 泊 3 日の広島旅行
学部生 ($n = 11$)	東インド会社の設立から解散まで, アメリカ合衆国の成り立ち, 幕末期のアジア史, 第二次世界大戦, フランス革命, 三国志	友だち 3〜4 人と 5 日間の冬スキー場への旅行, 友人 4 人と紅葉の時期に 2 泊 3 日で広島旅行, 友だち 5 人と春休みに 3 泊 4 日の沖縄旅行, 両親と 3 人で 2・3 月頃に 6 泊 7 日の京都旅行, 友だち 3〜4 人と年末に 1 週間位の北海道旅行

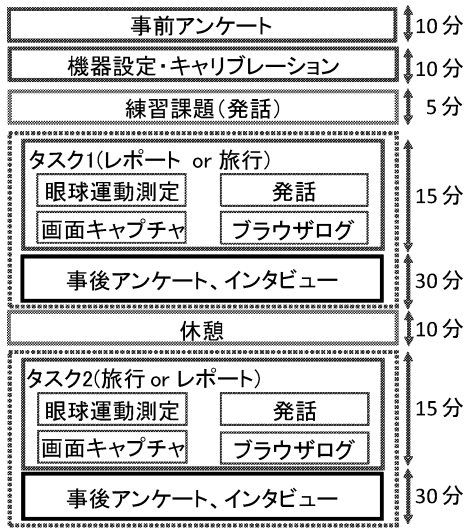


図 1　実験手続き

ンジンを用いた日常的な検索経験についての質問に回答した．発話練習を兼ねた練習課題に 5 分間取り組んだ後，2 つのタスクにそれぞれ 15 分間取り組んだ．タスクはそれぞれレポートタスクと旅行タスクで，順序による影響を考慮し，実験参加者によってタスクの遂行順序がランダムになるよう配置した．さらに，タスク遂行中の思考内容を明らかにするため，実験参加者には検索遂行中に発話するよう求めた．タスク終了後は，タスクについての困難度や満足度等を問う事後アンケートを実施した．アンケート終了後，タスクに取り組んでいる際の実験参加者の情報探索行動について，インタビューを実施した．インタビューでは，検索時の記憶想起を促すため，画面キャプチャー映像を参照しながら進めた．

検索遂行中の実験参加者の眼球の動きは，眼球運動測定装置（Voxer ST-600）によって計測した．実験には 19 インチ液晶モニタに対し，1024 × 786 の解像度に設定した Windows XP の PC を使用した．Web 閲覧用のブラウザには Firefox を使用した．ブラウザのブックマークには Google および Yahoo! Japan のトップページを追加し，日常での自然な探索と同様の状態となるよう，自由に探索を行うよう指示した．なお, 眼球運動データの分析のために, 原則としてブラウザは画面上にウィンドウサイズを最大化した状態で使用するよう促し，必要に応じてタブ機能を使うよう指示した．また，ブラウザのログは Slogger，コンピュータの画面は画面キャプチャーソフト HyperCam によって記録した．なお，ブラウザ環境および PC 操作環境において，日常的に使用しているものとの違いを実験参加者に尋ねたところ，数人に若干のとまどいを感じたとの返答もあったものの，おおむね軽微な影響にとどまるとの回答であった．

4　分析手法

本章では，前節で述べた探索実験の結果得られた行動データおよび眼球運動データに対する分析方法について述べる．

実験参加者のタスク遂行過程の分析に先立ち，眼球運動データとブラウザの閲覧行動ログをもとにして，探索過程の時系列上に対して，ページ種別，行動種別，ページ遷移，眼球運動注視箇所をそれぞれ人手でタグ付けした．タグ付けの詳細については分析結果とともに述べる．

タグ付けのためのツール COPATT（図 2）を開

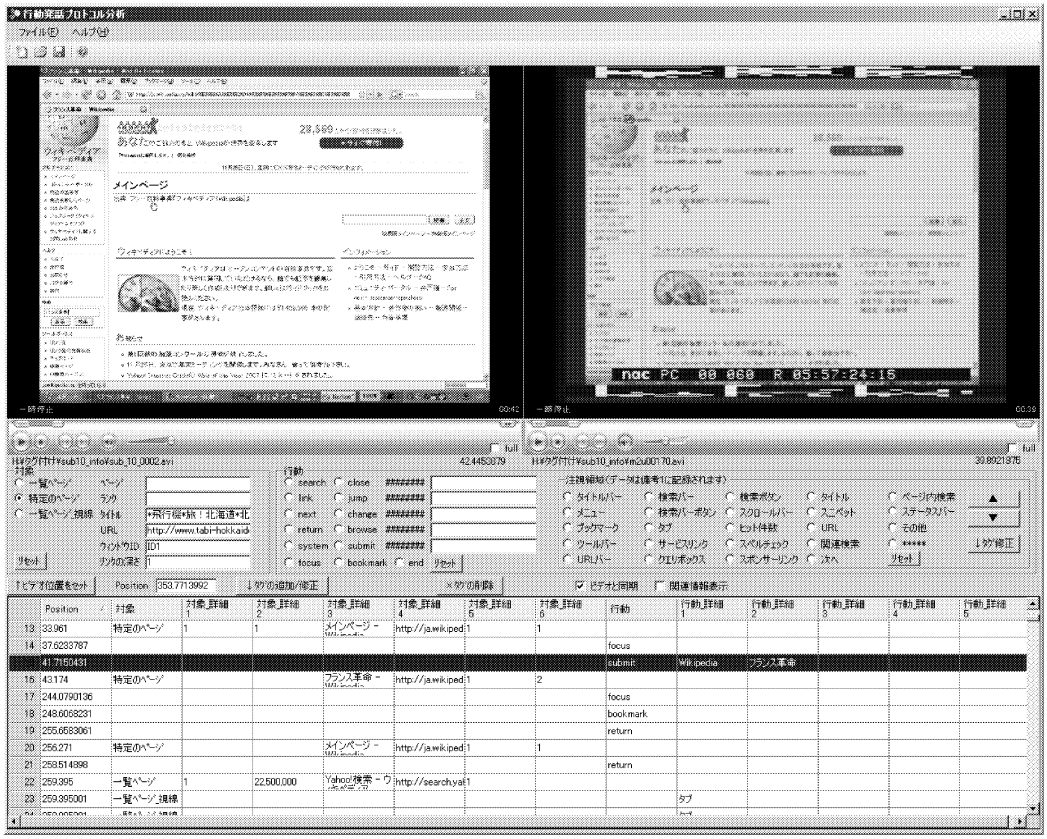


図 2　タグ付けツール：COPATT

スクリーンショットの左上に発話音声付きの画面キャプチャ映像，右上に眼球運動映像またはインタビュー映像（選択ボタンにより表示を切り替える）が表示されている．中央部がタグ付けのための入力インタフェース，下部がタグ付け結果となっている．本ツールでは，タグ付け結果は，左上の画面キャプチャ映像のタイムスタンプとともに保存されており，画面キャプチャ映像と下部に表示されているデータは同期して表示され，下部のタグ付け結果をクリックすることにより，同タイムスタンプの映像の閲覧が可能となっている．同様に，右上の眼球運動映像またはインタビュー映像も同様に同期が取られている．なお，タグ付けした結果は，XML ファイルとして出力される．

発し，ブラウザログと発話音声付きの画面キャプチャ映像，眼球運動映像，インタビュー映像に基づき，実験中に記録したデータに人手でタグ付けを行った．

タグ付けされた行動データと眼球運動データの分析を行い，タスク間，グループ間の比較を行った．なお，統計的な検定は，被験者内要因としてタスク（レポート/旅行）の 2 水準，被験者間要因としてグループ（大学院生/学部生）の 2 水準の 2 要因混合分散分析を使用した．また多重比較検定には LSD を使用した．検定の結果は，5% 水準を有意とし，10% 水準を有意傾向として示した．

探索行動データの分析手法は次のとおりである（4.1 節参照）．

- 実験参加者がサーチエンジンに発行したクエリ数
- 閲覧された Web ページをサーチエンジンの検索結果ページ（SERP）とその他のペー

ジ（non-SERP）の 2 つに分類した，ページの閲覧回数と閲覧時間
- 実験参加者の Web ブラウザ上の操作を 10 のカテゴリ Web 行動カテゴリに分類した行動数
- SERP ページからどれくらい離れてページを遷移したかを分析するための指標 Link Depth と Web 行動カテゴリを組み合わせた可視化結果と Link Depth ごとの閲覧回数

眼球運動の分析としては，

- SERP を 22 ブロック Lookzone に分割しブロックごとの注視回数を分析した結果
- SERP のランクの推移を示す指標 Scanpath とクリックしたランクを可視化した結果と Scanpath の分析結果
- SERP の注視ランクとクリックしたランクを分析した結果

の 3 つの観点からの分析結果を示す（4.2 節参照）．

なお，以下の各節の分析手法を用いた分析結果は，5 章における各節にそれぞれ対応させて報告する．

4.1 行動の分析

4.1.1 クエリの数

実験参加者によるサーチエンジンへのクエリの発行回数を分析する．両タスク遂行中の各実験参加者による発行クエリ異なり数をそれぞれ計数し，タスク単位・ユーザグループ単位での比較を行う．

4.1.2 ページ種別と閲覧回数，閲覧時間

閲覧ページは，サーチエンジンの検索結果一覧ページ（SERP: Search Engine Results Pages）と，それ以外の Web ページ（non-SERP: non Search Engine Results Pages）に分類した．ブラウザログと画面キャプチャー映像にもとづき，ページを表示した時間と種類のタグ付けを行なった．

タグ付けの結果から，SERP ページと non-SERP ページの閲覧回数や閲覧時間を分析した．

4.1.3 Web 行動カテゴリの分類と行動数

探索行動を分析するために，表 2 に示す 10 種類の Web 行動カテゴリを定義した．この定義は，齋藤，三輪 [36] が提案した WWW 情報探索における行動のカテゴリ分け 6 種（Search, Link, Next, Back, Jump, Browse）に，新たに 4 つのカテゴリ，Submit, Bookmark, Change, Close を追加したものである．

なお，ここで 4 つのカテゴリを追加した理由は，タスクの複雑さと，タスクの探索状況に応じた分析をおこなうためであった．すなわち，Submit 行動は動的サイト内での行動内容をとらえるために必要であり，Change および Close 行動は，複数ウィンドウおよびタブを使った閲覧行動の区切りを示すために必要である．また Bookmark 行動は，3.3 節で述べたように，探索タスクにおいて目標サイト発見を示すために利用したため，この行動を直接に分析できるようタグ付け対象として加えた．

ブラウザログと画面キャプチャー映像に基づき，行動の発生時刻と種類のタグ付けを行った．

4.1.4 Link Depth

サーチエンジンの検索結果一覧ページ（SERP）からどれくらいリンクをたどって Web 空間を移動したかを表す指標 Link Depth にもとづく分析を提案する．

サーチエンジンを対象とした Web 検索の研究では，SERP からのページ遷移情報である，クリックスルーデータを用いた研究が盛んに行われている [37][38][39]．一方，本研究のようにクライアントサイドログにもとづく分析が行える場合には，サーバサイドのクリックスルーログからだけでは捉えることのできない，サーチエンジンから検索結果を取得した後の探索行動を含めた全体の情報探索行動の特徴を明らかにすることができる．

指標 Link Depth は以下のように定義される：

- SERP ページからのリンクをたどって閲覧したページの Link Depth を 1 とする．
- そのページ中にあるリンクをたどったページは 2 とし，リンクをたどるごとに 1 増やす．
- ブラウザの Back ボタンで戻った場合は -1，Next ボタンで進んだ場合は，$+1$ とする．
- あらかじめブックマークしていたページや履歴をたどった場合は，ブックマーク追加時点もしくは履歴元ページと同じ Link Depth とする．

表 2　Web 行動カテゴリの定義

カテゴリ名	定義
Search	検索エンジンを使って検索する． クエリ入力終了後に，サーチエンジンの検索実行ボタンをクリックしたり，クエリの入力フォーム中で Enter キーを押したりなどして検索を実行すること．
Link	リンクをクリックする． ページ内リンクもこのカテゴリに入る．Ajax などで URL が変化せずとも，リンクに相当するアイコン等をクリックして閲覧内容が変わる場合はこのカテゴリとする．
Next	履歴のひとつ先へ進む． ブラウザの「進む」ボタンをクリックすること．
Back	履歴のひとつ前へ戻る． ブラウザの「戻る」ボタンをクリックすること．
Jump	履歴の 2 つ以上前または後ろに移動する． 履歴の直前に移動した場合は Back，履歴の直後に移動した場合は Next とし，それ以外を Jump とする．「履歴」メニューや URL アドレスのメニューを操作して 2 つ以上前の Web ページをクリックすること．ブックマークメニューから以前にブックマークしておいたページを選択すること，「進む」ボタンの右メニューから 2 つ以上離れたページを選択することなどである．
Browse	SERP ページの前または後ろの n 件目への SERP ページへ移動する． SERP ページの上部や下部にある「1 2 3 4 5 6 7 8 9 10 次へ」などとなっている別の SERP ページへ移動するためのリンクをクリックすること．
Submit	動的なページ内容を表示させるための行動． たとえば，路線検索サイトにおいて行き先等をメニュー選択した後，その路線検索結果を得るためにフォームなどの実行ボタンをクリックするなど．
Bookmark	閲覧ページをブックマークに追加する行動． ブックマークメニューから「このページをブックマーク」を選び，「完了」ボタンをクリックすることを指す．ブックマークに登録しておいたページに移動する行動はこのカテゴリではなく Jump カテゴリである．
Change	ウィンドウやタブを切り替える． 複数のウインドウやタブが開いている場合にのみ起こる行動である．他のウインドウやタブを表示させるために，ウインドウやタブの上部をクリックすること．
Close	ウィンドウやタグを閉じる． 複数のウインドウやタブが開いている場合にのみ起こる行動である．ウインドウやタブの右上にあるバツボタンをクリックすること．

注：「定義」列のブラウザの操作に関する説明は今回の実験に使用した Firefox における操作法による．なお，これらの操作は代表的な操作例を挙げたものである．

また，以下のページについては，Link Depth の対象外とした．

- 検索結果一覧ページ（SERP）
- サーチエンジンのトップページ
- 空白ページ （blank page）

図 3 は Link Depth の概念図である．最上段は Link Depth と対象外ページの名前をあらわしており，

- 「SE」はサーチエンジンのトップページをあらわし，（a）は SE である．
- 「SR」は検索結果一覧ページ（SERP）をあらわし，（b）は SR である．
- 「Blank」は空白ページをあらわし，（h）は空白ページである．

- 「1」は Link Depth 1 をあらわし，(c) と (e) の Link Depth は 1 である．
- 「2」は Link Depth 2 をあらわし，(d) と (f)，(i) の Link Depth は 2 である．

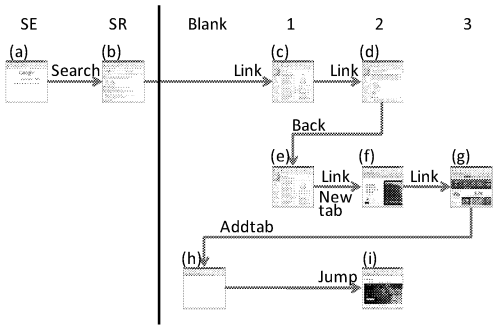


図 3　Link Depth の概念図

たとえば，実験参加者が Google のトップページ (a) で検索し，SERP ページ (b) が表示される．SERP ページからリンクをたどって表示された Web ページ (c) は，SR から最初にたどったページのため Link Depth は 1 となる．(c) からリンクをたどった (d) は，(c) の Link Depth に 1 を加えて 2 となる．(d) で back ボタンを押して表示した (e) は (d) の Link Depth2 から 1 を引いて，1 となる．(f) は新しいタブで開いているが，(e) からリンクをたどっているため，Link Depth は 1 を加えて 2 となる．(g) は (f) からリンクをたどっているため，1 を加えて 3 となる．(h) は新しくタブを開いて空白ページが表示されたため，Link Depth はつかない．(i) は以前にブックマークしていたページにたどっているので，ブックマークしたときに付いていた Link Depth となる．

4.2　眼球運動の分析

本節では眼球運動測定装置から得られた実験参加者の眼球運動データに基づく分析手法を示す．

4.2.1　SERP Lookzone 分類とその注視回数

ユーザが情報探索を行う画面上のブラウザ操作関連の要素および SERP ページコンテンツ内の共通要素を，図 4 に示す 22 項目の SERP Lookzone として定義し，注視点のタグ付けに使用した [6]．図 4 では SERP Lookzone の各ブロックの位置を Google の検索結果ページで示したが，これらの項目は今回の実験参加者が使用した他のサーチエンジン（Yahoo!Japan, MSN）でも同様に分類することができる．

次に実験時に記録した実験参加者の眼球運動データを元に，SERP の表示が開始された時刻を基点にして 0.5 秒間隔で画像を切り出した．その上で，抽出した画像に表示された注視点が Lookzone のいずれに当てはまるかを人手で確認しながらタグ付けした．以下での分析ではこの 0.5 秒間隔で抽出した注視点に対する分析を行った．

眼球運動測定装置を用いた分析では一般に，眼球運動により観察されたデータが一定の領域に一定時間（150～500msec 程度）停留した場合に，これを停留点として，当該領域への注視が行われたとすることが多い [40]．本稿での分析ではこれを用いる代わりに，0.5 秒おきの注視点を用いるにとどめた．これは，Web 情報探索行動が非常に迅速な情報の取捨選択をともなうものであり，かつ，眼球運動測定の対象となる画面のブラウザ上でのスクロールやページ遷移行動などによって，刺激提示画面上の X-Y 軸による画一的な分析になじまず，停留点の単位と画面上で素早く変化する注視領域の判定を行うことが困難であるために，このような分析を行った．検索結果ページのみを対象とすれば，停留点にかかわる分析も一定の範囲で可能であるが，自然な状況での情報探索を行うとの本研究の目的に沿わないものとなる懸念から，その方式は採用しなかった．停留点分析には概して専用ソフトウェアが必要であり，かつ，上記で述べたような制約もあるため採用しなかったが，一方で一定時間間隔による注視点に対するタグ付けは人手による労力を伴うものの，わかりやすい Lookzone が定義されてあれば比較的容易であり，高価な専用の分析ソフトウェアを使う必要もなく，安価なビデオツールの組み合わせで分析できる点に利点がある．

4.2.2　Scanpath

図 5 に，ブラウザスクリーンショット上に注視点座標を描画したものを示す．おおむね SERP ページでの注視点の遷移を反映したものとなっている．たとえば，図 5 はキーワード「三国志」で検索した SERP ページであるが，注視点は (1) 上部にある Yahoo カテゴリのリンク，(2) Yahoo オークション等の Yahoo 内で提供

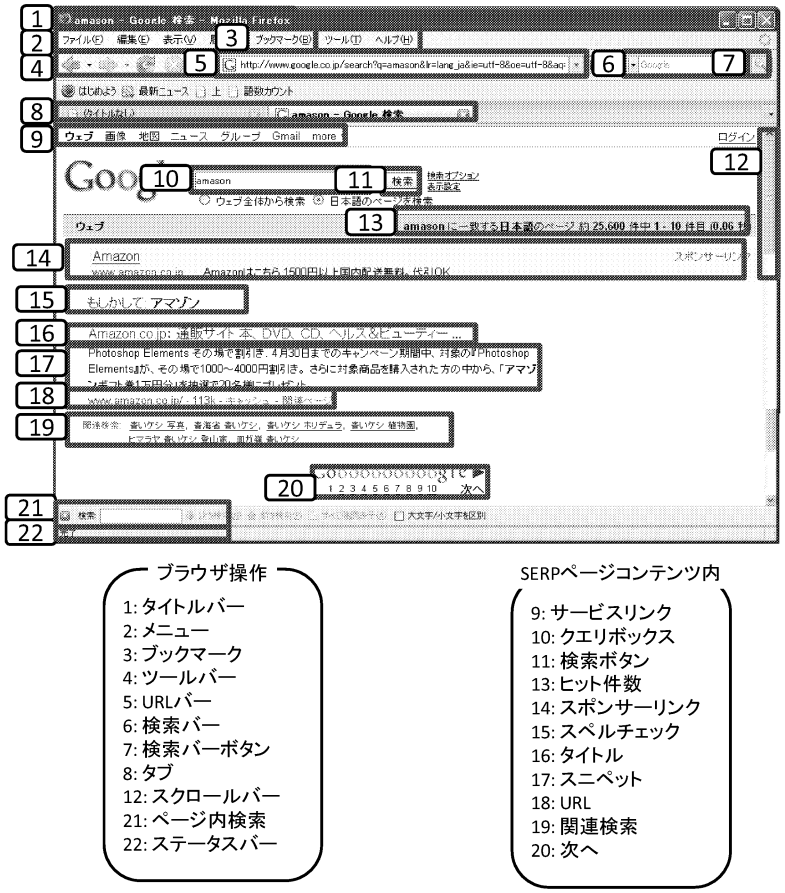


図 4　SERP Lookzone

ユーザが情報探索を行う画面上のブラウザ操作関連の要素および SERP ページコンテンツ内の共通要素を 22 項目の SERP Lookzone として定義し，各ブロックの位置を Google の検索結果ページで示した図である．各ブロックを四角で囲んで示している．各ブロックの左に各ブロックの ID を示している．各ブロックの ID は左上から順に通し番号となっている．

されているサービスへのリンクに目を向けたあと，(3〜6) クエリボックス下にいくつか表示されている関連検索キーワードに目を向け，そのあとで (7〜10) ランキングを上位の 2 件目・1 件目から眺め，つづいて (11〜13) 3・4・5 件目をざっと眺めて，再度，(14〜18) サービスリンク等に目を向けている．この後，実験参加者は Yahoo カテゴリの「中国文学 > 三國志」というリンクをクリックしている．これは，このページにおける滞在のうち最後のほうの注視点 17・18 の付近と一致するものとなっていることが分かる．

実験参加者が一覧ページ中のランキングを実際にどのように閲覧し，探索行動をたどったかを明らかにするため，Lorigo ら [33] の提案手法である，探索時の SERP ページ中のランキング部分の注視点 Scanpath により，その推移を分析する．Scanpath 分析では，注視ブロックのうちランキング部位である，スニペット・タイトル・URL (図 4 中の 16, 17, 18) に注視点があった場合，それぞれ該当ランクを注視していたも

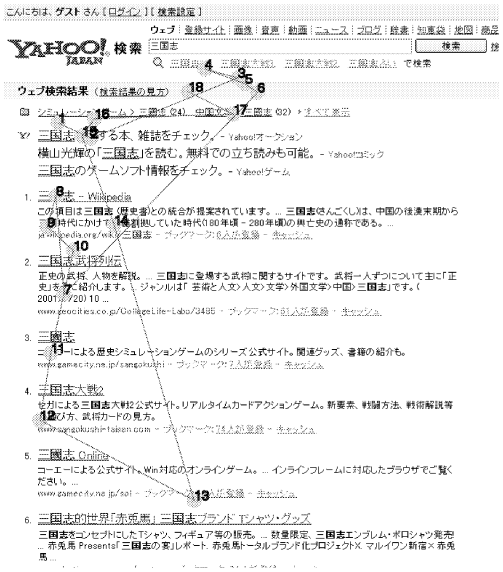


図 5　注視点座標のスクリーンショット中への図示例

図中の丸印は停留点を模式的に表したもの．丸印に示されている数字は停留点の推移を表しており，数字が増える方向にサッケードによって停留点が推移している様子を示している．

のとしてカウントし，そのランキング推移を系列として並べる．たとえば実験参加者の注視点が，ランキング中で，ランク 2, ランク 1, ランク 2, ランク 2, ランク 3 という順に移動していれば，「2-1-2-2-3」と表現する．この際，ランキング以外への注視などは系列から削除する．

さらに，圧縮 Scanpath として，同一ランキング注視が連続している系列を 1 つの注視としてカウントする方式も用いる．先の例を圧縮 Scanpath で表現すると，「2-1-2-3」となり，同一ランクへの注視の連続が圧縮できる．圧縮 Scanpath は，ランキング間の推移を比べるのに適している．

4.2.3　各文書ランクの注視回数とクリック回数

SERP ページにおいては「スニペット」,「タイトル」,「URL」の 3 つの Lookzone からなる検索結果文書の要約箇所がユーザの注目をもっとも集める領域となる．つづいて，これらの要約箇所の注視がそれぞれの文書ランク毎に異なるかを分析する．要約箇所の領域（前述の「タイトル」,「スニペット」,「URL」の 3 ブロック）の注視回数をまとめてランキングごとに分け，どのランクに対する注視が多いのかを示す．また，行動データから実際に各ランク文書をクリックしてたどったランクの情報を抽出し，注視点とクリックとの関係を検討する．

5　結果

本章では，4 章で述べた分析方法を，今回実施した探索実験結果の行動データおよび眼球運動データに対して適用した結果を述べる．なお，各節の結果に対する分析手法は 4 章の各節に述べたものに対応する．

5.1　行動の分析

5.1.1　クエリの数

図 6 は，実験参加者が使用したクエリのユニーククエリ数を示している．分散分析の結果，タスク・グループ間で有意な差は見られなかった．

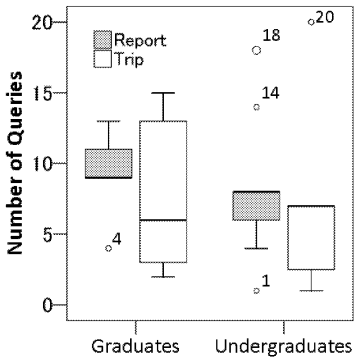
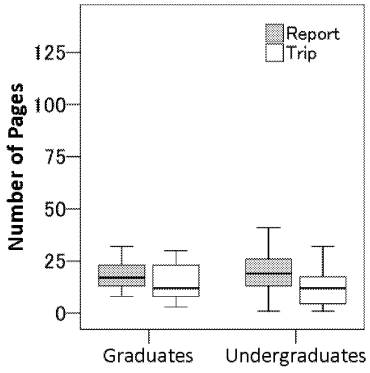


図 6　大学院生・学部生のタスクごとのクエリの数

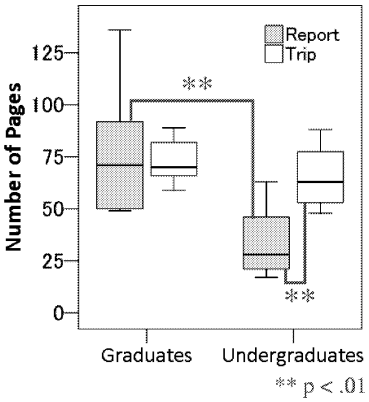
5.1.2　ページ種別と閲覧回数，閲覧時間

図 7 は，大学院生・学部生のタスクごとの SERP ページの閲覧回数と non-SERP ページの閲覧回数を示している．

閲覧回数を分析した結果，SERP ページではタスク・グループに有意な差は見られなかった．一方 non-SERP ページでは，交互作用が有意だった（$F(1,14)=10.75$, $p < .01$）．下位検定の結果，レポートタスクにおいてグループの差が見られ（$F(1,14)=12.47$, $p < .01$），大学院生は学部生よりも多く閲覧していた．また学部生においてタスクの差が見られ（$F(1,14)=14.73$,

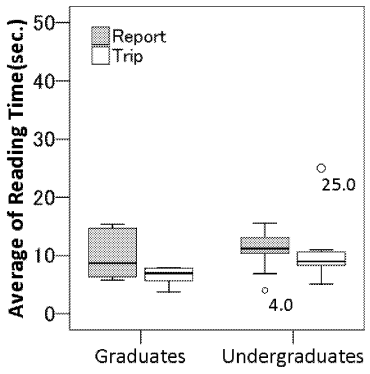


(a) SERP ページ

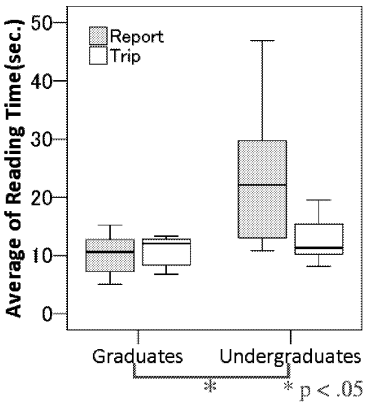


(b) non-SERP ページ

図 7　大学院生・学部生のタスクごとの SERP/non-SERP ページの閲覧回数



(a) SERP ページ



(b) non-SERP ページ

図 8　大学院生・学部生のタスクごとの SERP/non-SERP ページの平均閲覧時間

$p < .01$)，学部生はレポートタスクより旅行タスクでページを多く閲覧していた．

次にページ閲覧時間の分析を示す．図 8 は，大学院生・学部生のタスクごとの SERP ページの平均閲覧時間と non-SERP ページの平均閲覧時間を示している．分析の結果，SERP ページでは閲覧時間にタスク・グループに有意な差は見られなかった．一方 non-SERP ページでは，グループの主効果に有意差があり（$F(1,14)=5.75$, $p < .05$)，学部生は大学院生よりも閲覧時間が長かった．

分析の結果，SERP ページでは，閲覧回数・閲覧時間ともにタスク種別やユーザ特性の違いが見られなかった．一方 non-SERP ページでは，ユーザ特性の違いによって，タスクの影響が異なっていた．大学院生は，タスク種別の影響が見られなかった．それに対して，学部生はタスク種別の影響を受け，レポートタスクではページ閲覧回数が少なく，旅行タスクではページ閲覧回数に大学院生と差は見られなかった．閲覧時間の影響を受けたものと思われる．

5.1.3　Web 行動カテゴリの分類と行動数

表 3 は大学院生・学部生のタスクごとの行動数の平均と標準偏差（SD）を示している．分散

表 3　大学院生・学部生のタスクごとの行動数の比較

Web 行動カテゴリ	大学院生 (n=5)				学部生 (n=11)				グループ	タスク	交互作用
	レポート		旅行		レポート		旅行				
	平均	(SD)	平均	(SD)	平均	(SD)	平均	(SD)			
Search	9.20	(2.99)	7.80	(5.27)	8.00	(4.37)	6.27	(4.92)	n.s.	n.s.	n.s
Link	28.80	(7.28)	33.20	(8.57)	19.36	(6.26)	35.64	(8.65)	n.s.	**	*
Next	0.80	(0.75)	0.20	(0.40)	0.45	(0.78)	0.91	(1.08)	n.s.	n.s.	+
Back	10.40	(8.11)	10.80	(7.19)	17.45	(7.51)	22.27	(13.80)	+	n.s.	n.s.
Jump	2.20	(1.72)	3.40	(2.25)	2.64	(1.61)	2.64	(1.92)	n.s.	n.s.	n.s
Browse	0.80	(1.17)	0.60	(1.20)	1.82	(2.25)	0.18	(0.57)	n.s.	n.s.	n.s
Submit	7.60	(11.29)	4.60	(4.84)	1.27	(2.60)	3.00	(2.80)	+	n.s.	n.s.
Bookmark	8.00	(1.26)	8.00	(5.76)	4.55	(2.06)	4.55	(2.31)	*	n.s.	n.s.
Change	43.60	(23.59)	28.40	(17.85)	2.45	(5.37)	3.55	(3.23)	**	**	**
Close	4.20	(3.54)	6.00	(8.79)	0.36	(0.64)	2.36	(1.77)	+	+	n.s.

凡例：

**		有意差 ($p < .01$) があったことを示す.
*		有意差 ($p < .05$) があったことを示す.
+		有意傾向 ($p < .10$) があったことを示す.
n.s.		有意差，有意傾向ともになかったことを示す.
123	（二重下線）	交互作用の下位検定の結果，グループ間に有意差 ($p < .01$) があったことを示す.
123	（下線）	交互作用の下位検定の結果，グループ間に有意差 ($p < .05$) があったことを示す.
123	（あみかけ）	交互作用の下位検定の結果，タスク間に有意差 ($p < .01$) があったことを示す.

分析の結果，Search，Next，Jump，Browse においては有意な差は見られなかった.

Link においては，交互作用が認められた ($F(1,14)=4.62$, $p < .05$)．下位検定の結果，学部生のみタスク間の違いが認められ，レポートタスクよりも旅行タスクの方が Link 行動が有意に多かった ($F(1,14)=17.37$, $p < .01$)．グループ間の違いはレポートタスクにおいてのみ認められ，学部生より大学院生の方が有意に Link 行動が多かった ($F(1,14)=6.16$, $p < .05$)．Back においては，グループ間の違いが認められ，大学院生より学部生の方が Back 行動が多い傾向が見られた ($F(1,14)=3.63$, $p < .10$)．一方，Submit においては，学部生より大学院生の方が多く Submit 行動を行う傾向が見られた ($F(1,14)=3.79$, $p < .10$)．

Bookmark においては，グループ間での差異が認められ，学部生より大学院生の方が有意に多かった ($F(1,14)=6.86$, $p < .05$)．

Change においては，交互作用が認められた ($F(1,14)=12.10$, $p < .01$)．下位検定の結果，グループ間の差異が認められ，学部生より大学院生の方が Change 行動が有意に多い結果となった（レポートタスク: $F(1,14)=26.28$, $p < .01$; 旅行タスク: $F(1,14)=17.41$, $p < .01$)．加えて，大学院生においては，タスク間の違いが認められ，旅行タスク よりレポートタスク の方が Change 行動が有意に多い結果となった ($F(1,14)=21.06$, $p < .01$)．

Close 行動は，学部生より大学院生の方が多い傾向が見られた ($F(1,14)=3.50$, $p < .10$)．さらにどちらのグループでもレポートタスクより旅行タスクにおいて Close が多い傾向にあった ($F(1,14)=3.33$, $p < .10$)．

行動カテゴリの分析におけるグループとタスク間の差異についてまとめる．大学院生は複数のウィンドウやタブを切り替える行動である Change やウィンドウやタブを閉じる行動である Close が多く見られた．この結果は，大学院生が複数のページを同時に開き，切り替え行動を頻繁に行いながら並列的に探索をしていたことを示している.

5.1.4 Link Depth

■Link Depth の可視化　時系列に沿った Link Depth の可視化を行った．図 9，10 に，それぞれ実験参加者 1 名分の探索行動全体を可視化した例を示す．可視化結果より，リンクをたどる行動とともに，検索行動とページ閲覧行動の全体像を一目で明らかにでき，レポートタスク, 旅行タスクの複数の探索系列における違いも並べて眺めることで容易に把握できる．

図 9 の実験参加者（学部生）の場合は，レポートタスクでは，閲覧したほとんどのページは，サーチエンジンの検索結果ページやサーチエンジンの結果から直接リンクでたどったページであり，検索結果ページにおいて検索キーワードを何度も書き換えながらの探索を行っていることがわかる．一方，旅行タスクではサーチエンジンでの検索は 2 回実行しているのみで，特定ページ中の多くのリンクをたどって Web 空間を移動していることがわかる．

図 10 の実験参加者（大学院生）の場合は，どちらのタスクも多くのキーワードを使って検索していることや，サーチエンジンの結果から近いページも遠いページもどちらとも多く見ていることがわかる．

可視化した結果，図 9 で示した以外の学部生についても, レポートタスクでは, SERP ページおよび SERP ページからたどった深さ 1 のページを多く閲覧している傾向があることが判った．また，学部生，大学院生ともに，旅行タスクでは，SERP ページから離れて，多くのリンク先をたどる行動を行っている傾向が示唆された．

■Link Depth ごとのページ閲覧回数　図 11 に，各タスク遂行中の Link Depth の分布図を示す．Link Depth の最大値をタスクごとに比べたところ，旅行タスクのほうが Link Depth の最大値が大きく，より深くリンクをたどっていることがわかった．Link Depth の最大値と平均値の分散分析の結果，タスクの主効果に有意な差があった（最大値：$F(1,14)=6.48, p<.05$；平均値：$F(1,14)=7.68, p<.05$）．

図 12 は，Link Depth 1 のページと 2 以上のページの閲覧回数を示した図である．Link Depth 1 のページはサーチエンジンの検索結果一覧ページから直接たどれるページであり, Link Depth2 以上のページはさらに深くリンクをたどったページである．

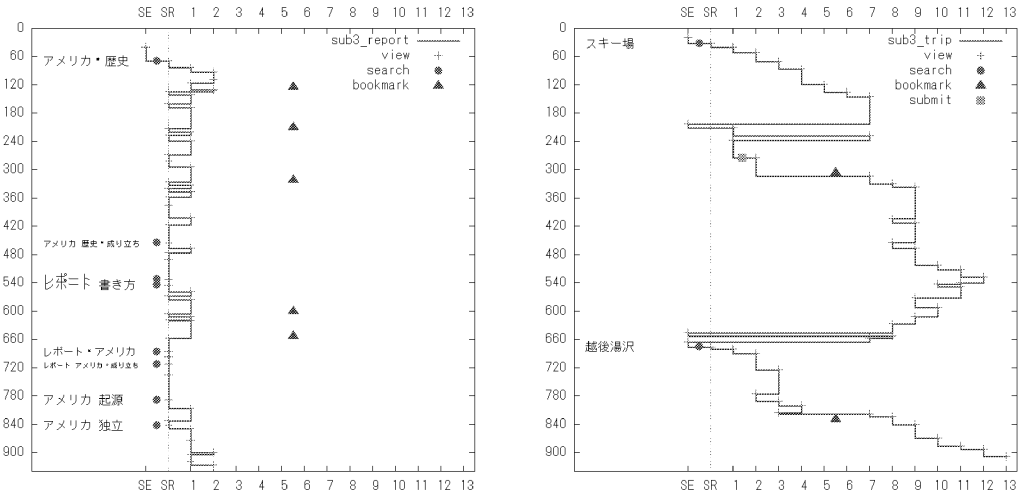
Link Depth が 1 のページと 2 以上のページの閲覧回数を分散分析した結果，Link Depth が 1 のページでは，グループの主効果に有意差があり（$F(1,14)=10.15, p<.01$），大学院生は，学部生より多く Link Depth が 1 のページを閲覧していた．タスクの主効果に有意傾向も見られ（$F(1,14)=3.68, p<.10$），旅行タスクよりレポートタスクのほうが多く Link Depth が 1 のページを閲覧していた．

Link Depth が 2 以上のページの閲覧回数では，グループの主効果に有意差があり（$F(1,14)=5.16, p<.05$），大学院生は，学部生より多く Link Depth が 2 以上のページを閲覧していた．タスクの主効果にも有意差があり（$F(1,14)=7.40, p<.05$），レポートタスクより旅行タスクのほうが多く Link Depth が 2 以上のページを閲覧していた（交互作用に有意傾向がでており（$F(1,14)=3.24, p<.10$），レポートタスクのグループ（$F(1,14)=6.04, p<.05$）と学部生のタスク（$F(1,14)=10.22, p<.01$）に有意差があった）．

以上の Link Depth による分析結果は，先述の図 7（b）で示された non-SERP ページにおけるタスク間およびグループ間の差異を詳細に説明したものとしてとらえることができる．つまり，学部生のレポートタスクにおいてページ閲覧回数が同じ学部生の旅行タスクよりも少ない点については，Link Depth ごとの分析結果から，Link Depth≧2 のページ閲覧回数に起因すると思われる．同様に，学部生のレポートタスクでの閲覧回数が大学院生のレポートタスクよりも少ない結果は，Link Depth の深さによらず，一貫した傾向を示していることが確認できる．

5.2 眼球運動の分析

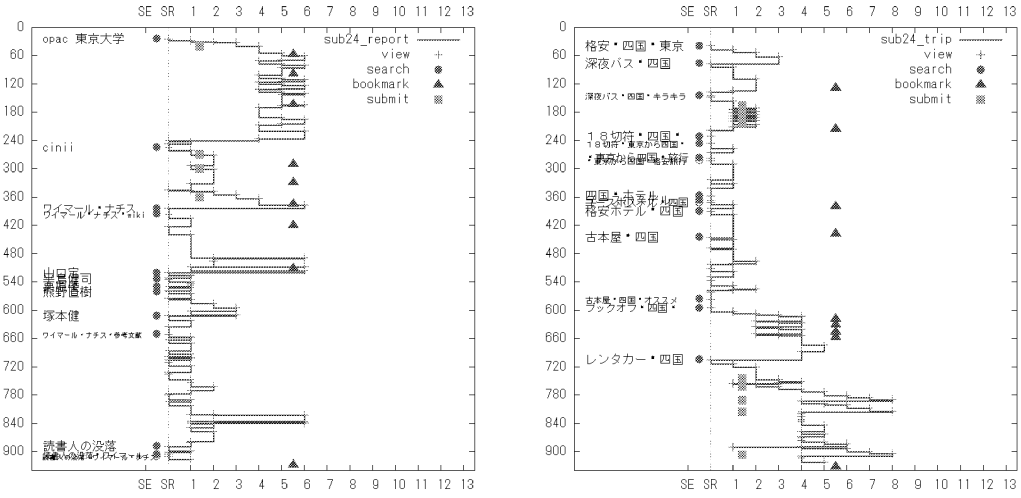
本節では眼球運動測定装置から得られた実験参加者の眼球運動データに基づく分析を示す．分析対象は，前述した実験において使用された，Google，Yahoo! Japan，MSN の 3 つの一般的なサーチエンジンにおける，SERP 上での視線注視を示すデータに基づく結果である．なお，実験参加者のうち学部生 2 名は，測定装置の環境において眼球運動データが得られなかったため，分析から除外した．



(a) レポートタスク　　(b) 旅行タスク

図 9　Link Depth 可視化の例（学部生）; 実験参加者 1 人分の行動とページ移動の可視化

左はレポートタスク遂行中の行動であり，右は旅行タスク遂行中の行動に基づく可視化である．X 軸は Link Depth の深さを示し，Y 軸は探索課題遂行中の経過秒数を示す．X 軸上の SE, SR はそれぞれサーチエンジントップページおよび SERP ページの閲覧を示している．検索クエリの実行（「Web 行動カテゴリ」の Search）には図中の丸点が対応し，SE の左側にある文字列が，その時点で利用者が入力した検索クエリをあらわす．また，図中の三角点はその時点で閲覧ページをブックマークに追加したこと（「Web 行動カテゴリ」の Bookmark）をあらわし，四角点はフォームボタンをクリックしたり，入力フォームでエンターキーを押したこと（「Web 行動カテゴリ」の Submit）を示している．



(a) レポートタスク　　(b) 旅行タスク

図 10　Link Depth 可視化の例（大学院生）; 実験参加者 1 人分の行動とページ移動の可視化

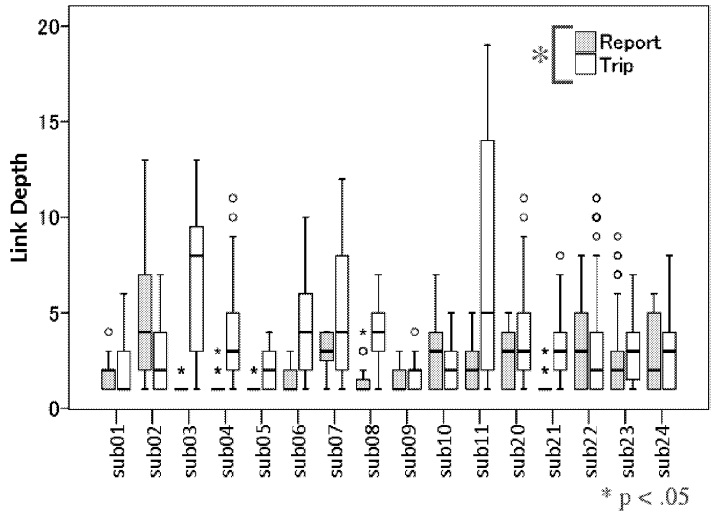
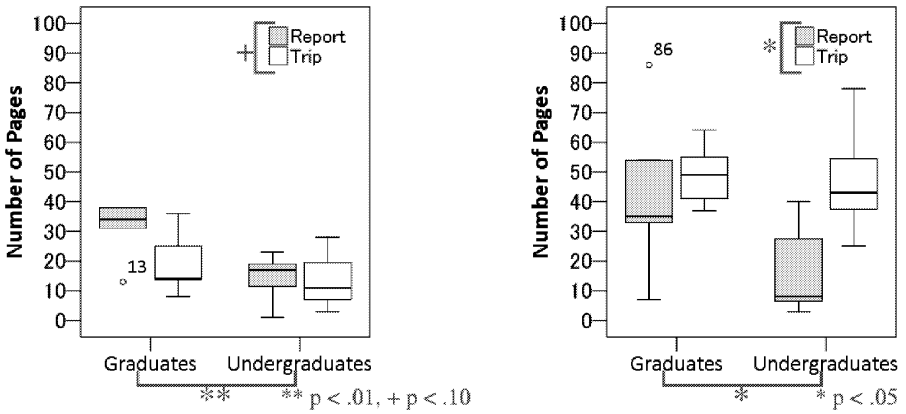


図 11　実験参加者ごとの Link Depth の分布図（学部生：sub01～sub11，大学院生：sub20～sub24）



(a) Link Depth=1　　(b) Link Depth≧2

図 12　大学院生・学部生のタスクごとの Link Depth 1 のページと 2 以上のページの閲覧回数

5.2.1　SERP Lookzone 分類とその注視回数

表 4 に一タスクあたりの上記の注視ブロックとして定義した各 Lookzone の注視回数平均と標準偏差を示す．Lookzone のうち，実際のユーザ実験における注視がほとんどなく，一タスクあたりの平均注視回数が 1 未満のブロックは，ブラウザ操作領域で 4 つ（「URL バー」，「スクロールバー」，「検索バーボタン」，「ページ内検索」）あり，SERP ページコンテンツ内要素で 3 つ（「次へ」，「ヒット件数」，「スペルチェック」）あった．つまり，たとえば，SERP 閲覧中の「スクロールバー」や「URL バー」への注視はほとんど起きなかったことが分かる．

表 4 からも明らかなように，Lookzone として設定したブロックのうち多くは注視回数そのものが少なかったり，全く注視を受けないブロックも多い．一方で，「タイトル」，「スニペット」，「URL」といった SERP において中心となる検索結果文書の情報に関するブロックに注視が集中していることが分かる．

表 4　一タスクあたりの各 Lookzone への平均注視回数

	ID	Lookzone	大学院生 (n=5)		学部生 (n=9)		グループ	タスク	交互作用
			レポート 平均 (SD)	旅行 平均 (SD)	レポート 平均 (SD)	旅行 平均 (SD)			
ブラウザ操作	1	タイトルバー	0.40 (0.80)	0.80 (0.98)	3.78 (6.81)	1.00 (1.56)	–	–	–
	2	メニュー	1.80 (3.12)	0.00 (0.00)	0.22 (0.42)	0.11 (0.31)	n.s.	n.s.	n.s.
	3	ブックマーク	0.00 (0.00)	0.20 (0.40)	4.22 (5.90)	0.00 (0.00)	n.s.	n.s.	n.s.
	4	ツールバー	0.40 (0.80)	0.40 (0.80)	1.33 (1.63)	1.22 (1.40)	n.s.	n.s.	n.s.
	5	URL バー	0.40 (0.49)	0.00 (0.00)	0.56 (1.07)	0.22 (0.42)	–	–	–
	6	検索バー	6.40 (7.50)	4.00 (7.04)	0.00 (0.00)	0.00 (0.00)	+	n.s.	n.s.
	7	検索バーボタン	0.40 (0.49)	0.20 (0.40)	0.00 (0.00)	0.00 (0.00)	–	–	–
	8	タブ	12.00 (14.13)	6.00 (6.63)	8.11 (9.81)	9.22 (17.94)	n.s.	n.s.	n.s.
	12	スクロールバー	0.60 (0.80)	0.00 (0.00)	0.11 (0.31)	0.00 (0.00)	n.s.	*	n.s.
	21	ページ内検索	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	–	–	–
	22	ステータスバー	0.00 (0.00)	0.00 (0.00)	1.78 (3.39)	0.00 (0.00)	n.s.	n.s.	n.s.
SERP ページコンテンツ内	9	サービスリンク	2.40 (2.06)	2.20 (2.14)	17.67 (23.44)	5.11 (9.33)	n.s.	n.s.	n.s.
	10	クエリボックス	5.60 (4.36)	3.00 (4.65)	36.89 (36.71)	12.56 (11.93)	+	n.s.	n.s.
	11	検索ボタン	0.00 (0.00)	0.20 (0.40)	0.89 (1.10)	0.67 (0.82)	+	n.s.	n.s.
	13	ヒット件数	0.00 (0.00)	0.60 (0.80)	0.44 (0.96)	0.00 (0.00)	–	–	–
	14	スポンサーリンク	0.00 (0.00)	11.40 (13.99)	6.67 (7.85)	12.44 (9.93)	n.s.	*	n.s.
	15	スペルチェック	0.00 (0.00)	0.20 (0.40)	0.00 (0.00)	0.00 (0.00)	–	–	–
	16	タイトル	41.20 (26.80)	39.20 (40.82)	59.67 (38.92)	42.11 (34.19)	n.s.	n.s.	n.s.
	17	スニペット	74.80 (42.56)	28.40 (28.00)	91.11 (55.59)	37.00 (32.84)	n.s.	**	n.s.
	18	URL	18.00 (9.21)	12.40 (11.83)	40.89 (34.27)	15.56 (11.35)	n.s.	n.s.	n.s.
	19	関連検索	1.20 (1.94)	1.20 (1.17)	3.00 (4.03)	2.56 (4.11)	n.s.	n.s.	n.s.
	20	次へ	1.00 (1.10)	1.00 (2.00)	1.00 (1.89)	0.78 (1.03)	–	–	–
		その他	21.60 (14.47)	17.00 (8.60)	52.89 (53.43)	18.89 (14.51)	–	–	–
		アイマーク不明	15.00 (10.94)	7.20 (4.71)	83.44 (73.10)	70.78 (101.06)	–	–	–

凡例：
** 有意差 ($p < .01$) があったことを示す．
* 有意差 ($p < .05$) があったことを示す．
+ 有意傾向 ($p < .10$) があったことを示す．
n.s. 有意差，有意傾向ともになかったことを示す．
– 分散分析の対象から除外したことを示す．

表 4 に示された各 Lookzone の注視回数に対して，分散分析を行った結果を以下に示す．なお，平均注視回数が 1 未満となった 7 ブロックについては，分散分析の対象から除外した．

分散分析の結果，複数のブロックでグループ間の違いが明らかになった．クエリボックスと検索ボタンでは，学部生の方が注視点が多い傾向が見られた（クエリボックス:$F(1,12)=3.84$, $p < .10$, 検索ボタン:$F(1,12)=4.72$, $p < .10$）．検索バーと検索バーボタンでは，大学院生の方が注視点が多い傾向が見られた（検索バー:$F(1,12)=4.71$, $p < .10$，検索バーボタン $F(1,12)=4.34$, $p < .10$）．同様に，複数のブロックでタスク間の違いが明らかになった．スクロールバーとスニペットでは，旅行タスクよりレポートタスクの方が注視点が有意に多かった（スクロールバー:$F(1,12)=4.77$, $p < .05$, スニペット:$F(1,12)=9.58$, $p < .01$）．またスポンサーリンクでは，レポートタスクより

旅行タスク の方が注視点が有意に多かった（$F(1,12)=6.15$, $p < .05$）．これらの結果から，レポートタスク では，実験参加者は SERP ページの下まで閲覧し，またページの概要であるスニペットを吟味していたことを示唆している．対して，旅行タスク では，実験参加者はスポンサーリンクを重視していたことを示唆している．

5.2.2 Scanpath

■Scanpath 長 表 5 に，Scanpath の平均の長さを示す．Scanpath と圧縮 Scanpath に対して分散分析を行なったところ，SERP ごとの Scanpath の長さの平均とクエリごとの長さの平均ともに全てにおいてタスクの主効果において有意傾向を示し，旅行タスクよりレポートタスクのほうが Scanpath が長い傾向が示された（SERP 単位の Scanpath: $F(1,12)=4.03$, $p < .10$; SERP 単位の圧縮 Scanpath: $F(1,12)=4.09$, $p < .10$; クエリ単位の Scanpath: $F(1,12)=3.84$, $p < .10$; クエリ単位の圧縮 Scanpath: $F(1,12)=3.58$, $p < .10$）．

■Scanpath の可視化 実験参加者がどのような Scanpath を描いたかを比べやすくするために，Lorigo ら [29] が行った可視化手法を参考に，Scanpath の可視化を行った．図 13 に，1 人の実験参加者が行ったレポートタスク全体の Scanpath を可視化したものを示す．この図を見ることにより，ランクを見た中のうち，31.1% はランク 1 位を見たことや，上位のものほど多く見ていることがわかる．また，発行したクエリや SERP 内の各ランクのタイトルやスニペットをどういう順でどのくらいの頻度で見たかがわかる．たとえば，最初のクエリ：「ウィキペディア」は，ランク 2 位を最初にみて，次にランク 1 位を見てこのクエリの検索を終了している．また，どのランクを見たあとに，リンクのクリックをしたかもわかり，たとえば，クエリ：「東京工業大学」は，最初に 1 位のランクをみて，そのままランク 1 位の領域をクリックしたことがわかる．

図 13 のように可視化することによって，1 位から 3 位あたりをよく見ていることなどや，9 位以降まで進んだクエリは 1 つしかないといった，リンクされたランクと SERP 上での注視箇所がたどった軌跡などを概観できる．

5.2.3 各文書ランクの注視回数とクリック回数

表 4 にも見られたとおり，SERP ページにおいては「スニペット」,「タイトル」,「URL」の 3 つの Lookzone からなる検索結果文書の要約箇所がユーザの注目をもっとも集める領域となる．

図 14 は，大学院生と学部生の各ランクごとのクリック数の割合と注視点の割合を示している．図をみると，大学院生も学部生もランク 1 のクリックおよび注視点の割合が最も高く，ランクが下がるにつれて注視およびクリックが減少する傾向を示していることが分かる．大学院生はどちらのタスクでもクリックや注視点の割合にさほど違いが見られないが，学部生は上位のランクにおけるクリックの割合がタスクによって異なる傾向が見られる．

注視点について各ランクで 2 要因分散分析を行った結果，ランク 4 とランク 7 においてタスクの主効果が有意であった（ランク 4: $F(1,12)=5.13$, $p < .05$, ランク 7: $F(1,12)=7.28$, $p < .05$）．この結果は，大学院生も学部生も，レポートタスクにおいて下位の

表 5 大学院生・学部生のタスクごとの Scanpath の平均値の比較

カウント単位	Scanpath の種類	大学院生 (n=5)		学部生 (n=9)		グループ	タスク	交互作用
		レポート 平均 (SD)	旅行 平均 (SD)	レポート 平均 (SD)	旅行 平均 (SD)			
SERP	Scanpath	7.25 (1.99)	3.97 (2.01)	8.55 (3.61)	6.8 (3.32)	n.s.	+	n.s.
SERP	圧縮 Scanpath	3.8 (1.35)	2.27 (1.15)	4.71 (2.00)	3.23 (1.70)	n.s.	+	n.s.
クエリ	Scanpath	14.92 (4.53)	8.47 (4.04)	21.66 (12.28)	13.7 (8.66)	n.s.	+	n.s.
クエリ	圧縮 Scanpath	7.61 (1.91)	4.82 (2.31)	11.92 (7.34)	6.67 (4.13)	n.s.	+	n.s.

凡例：
+ 有意傾向 ($p < .10$) があったことを示す．
n.s. 有意差，有意傾向ともになかったことを示す．

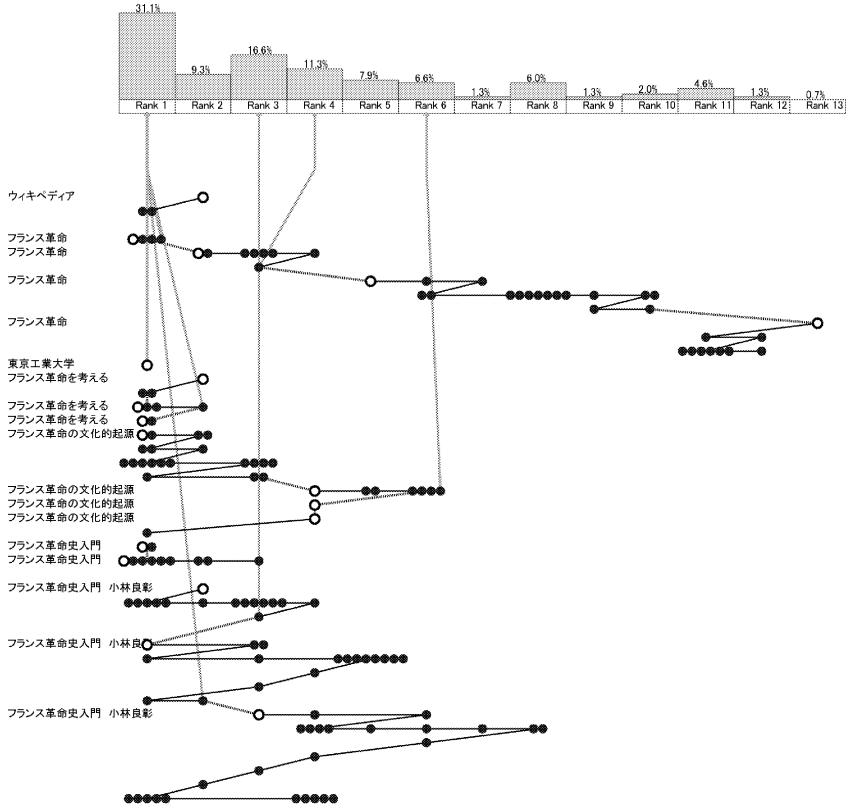
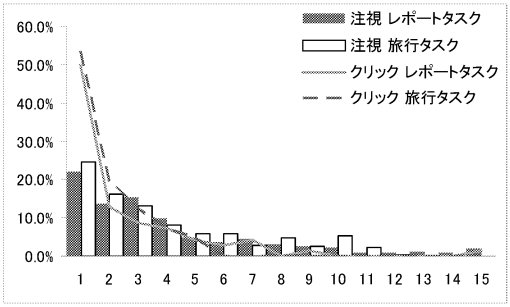


図 13　Scanpath を可視化した例 （sub10 レポートタスク）

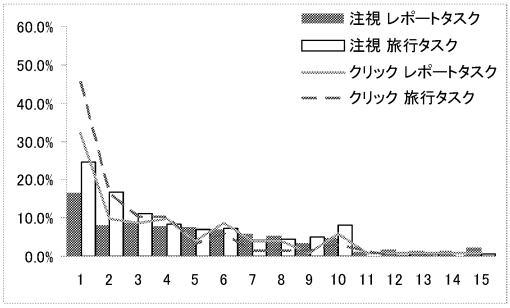
1 人の被験者が行ったレポートタスク全体の Scanpath を可視化した図である．上部の四角での “Rank n” ($n = 1, ..., 13$) は SERP 上での当該ランク n 位の領域を示している．“Rank n” 上部の棒グラフは，タスク遂行全体のうちランク領域を注視した数を分母として，それぞれのランクを見た割合を示している．図中の小さな円ノードは注視点を表しており，丸の数は注視点の数に対応することとなる．各ランク下に配置された丸ノードは，その当該ランク領域に注視点があったことを示す．白抜きの丸は各クエリにおける最初の注視点をあらわす．注視点の遷移順に丸ノードを線で結んでおり，クエリごとに分割している．実線はひとつの SERP 検索結果ページ上で推移した注視点であることを示し，点線の箇所は SERP からの一時的な離脱，つまり被験者がリンク行動などをおこなって別のページに移ったのちに再び検索結果ページに戻ってきた場合の注視点の推移であることを示している．発行されたクエリの内容は図の左側に示している．丸ノードから “Rank n” ボックスへの上方向の矢印は，その注視点直後に当該ランク文書へのリンク行動が行われたことを示している．

ランクまで閲覧していることを示している．

次にクリック数について各ランクで分散分析を行った結果，ランク 2, 7, 8, 10 においてグループ・タスクによる違いが見られた．まずランク 2 では，大学院生は学部生よりもランク 2 を多く選択する傾向が見られた ($F(1,12)=3.81$, $p < .10$)．ランク 7 では，どちらのグループも，旅行タスクよりレポートタスクにおいてランク 7 を多く選択していた ($F(1,12)=6.81$, $p < .05$)．またランク 8 およびランク 10 では，2 つのタスクにおいて大学院生がこれらのランクをクリックすることは無かったのに対し，学部生グループにはこれらのランクの選択行動が見られた (ランク 8: $F(1,12)=3.65$, $p < .10$, ラ



(a) 大学院生



(b) 学部生

図 14　各ランクのクリックと注視点の割合

ンク 10: $F(1,12)=4.12, p < .10$)．これらの結果は，大学院生は検索結果のランクが高いページを選択する傾向があるのに対し，学部生はよりランク下位のページも選択する傾向にあることを示している．

6　考察

前章で述べた分析結果を通じて観察された，タスク間およびユーザグループ間での探索行動における差異を表 6 にまとめて示す．

6.1　ユーザ特性が探索行動に与える影響

今回の実験で参加者として設定したユーザ特性は，図書館情報学を専攻とする大学院生と，学部生の 2 種類であった．本節では，この 2 種類の特性を持つユーザ間で Web 情報探索行動がどのように異なるかを示す．

ユーザ特性に関わる第一の特徴は，ページ閲覧行動にあたって，大学院生のほうが学生よりも多くのページを閲覧しているという点にある．この傾向はとりわけ non-SERP ページの閲覧で特徴的であり，Link Depth の分析でも，その深さを問わず，大学院生が一貫して学部生よりも多くのページを閲覧していることがわかる．その結果，一ページあたりの閲覧時間は全体として，学部生のほうが長くページを読んでいる．まとめると，大学院生は全般にすばやく効率的なページ閲覧を行っており，学部生はそれに比べると比較的じっくりとページ内容を読み込みながら判断を行っている．

さらに，この傾向はレポートタスクで顕著であり，学部生のレポート選択テーマに応じたレポート執筆計画の詳細化にあたって，学部生が個々のページを読みながらテーマの細分化を果たそうとする一方で，大学院生はレポートのテーマ決めから情報収集までを効率的におこなおうとする傾向にあることが示唆される．これらのとくにレポートタスクにおける違いは，当初設定した Web 探索や検索スキルの違いを超えて，レポートや論文執筆の経験の違いが影響した可能性も高い．たとえば，Kim ら [15] も，このようなレポートのための情報収集プロセスの効率化は，レポート執筆のような学術的な活動に対する問題解決経験の差異としてとらえうるのではないかとの仮説を示唆している．また，Web 行動カテゴリによる分析において，大学院生による Bookmark の付与が学部生を上回る点などから，効率的な探索プロセスの結果として，有用なページを多く発見していったものと考えられる．

もうひとつのユーザ特性による差異として，ブラウザ機能の一部において，使い方に違いがあった．まず，Lookzone 分析の結果から，サーチエンジンの検索実行方法として，Web ページ上のクエリボックスと検索ボタンを使うか，ブラウザ上に用意されている検索バーを用いるかが，ユーザ特性の違いとして示されている．大学院生はブラウザ上の検索バーを用い，一方で学部生はサーチエンジンの通常のクエリボックスを使っている．たとえば，サーチエンジンの利用の際に日頃から検索バーを通じた検索機能を利用していれば，自らのタスク遂行にあたって，わざわざサーチエンジンへページを移らずともその場で検索でき，時間短縮ができると考えたものと思われる．これは，大学院生が効率的なブラウジングを指向している，との前述の考察を支持するものと考えられる．

表 6　タスク間，実験参加者グループ間での探索行動における差異

分析手法	タスク間での差異		グループ間での差異	
クエリの数	―――	―――	―――	―――
ページの閲覧回数	学部生の non-SERP	レポート < 旅行 **	レポートの non-SERP	大学院生 > 学部生 **
ページの平均閲覧時間	―――	―――	non-SERP	大学院生 < 学部生 *
Web 行動カテゴリ	学部生の Link	レポート < 旅行 **	Change	大学院生 > 学部生 **
	大学院生の Change	レポート > 旅行 **	Bookmark	大学院生 > 学部生 *
	Close	レポート < 旅行 +	レポートの Link	大学院生 > 学部生 *
			Back	大学院生 < 学部生 +
			Submit	大学院生 > 学部生 +
			Close	大学院生 > 学部生 +
Link Depth	Link Depth≧2 の閲覧ページ数	レポート < 旅行 *	Link Depth≧2 の閲覧ページ数	大学院生 > 学部生 *
	Link Depth=1 の閲覧ページ数	レポート > 旅行 +	Link Depth=1 の閲覧ページ数	大学院生 > 学部生 *
Lookzone	スニペット	レポート > 旅行 **	クエリボックス	大学院生 < 学部生 +
	スポンサーリンク	レポート < 旅行 *	検索バー	大学院生 > 学部生 +
	スクロールバー	レポート > 旅行 *	検索ボタン	大学院生 < 学部生 +
Scanpath	Scanpath の長さ	レポート > 旅行 +	―――	―――
各文書ランクの注視回数	ランク 4	レポート > 旅行 *	―――	―――
	ランク 7	レポート > 旅行 *		
各文書ランクのクリック回数	ランク 7	レポート > 旅行 *	ランク 2	大学院生 > 学部生 +
			ランク 8	大学院生 < 学部生 +
			ランク 10	大学院生 < 学部生 +

凡例：
** 有意差（$p < .01$）があったことを示す．
* 有意差（$p < .05$）があったことを示す．
+ 有意傾向（$p < .10$）があったことを示す．

また，もうひとつのブラウザ機能の使い方の違いとして，ブラウザ上のタブ機能の使用もユーザ特性による違いとして挙げられる．Web 行動カテゴリの Change 行動数の違いとして示されているとおり，大学院生は学部生よりも頻繁にタブ機能を使っていた．このタブ機能における差異は，タブブラウジング機能により可能となる，個々のタブに複数の情報をまとめて表示して，より並列的な探索や迅速なページ内容の確認を目指しているものと考えられる．

また，タブ機能の使い方の違いにおいては，ブラウザ機能に対する習熟度の差を直接的な要因と見ることもできる．事前アンケートの結果によれば，大学院生と学部生は，普段使用している Web ブラウザに違いがあった．学部生は 1 名を除き，Windows PC にインストールされているデフォルトのブラウザである Internet Explorer を使用していた（Internet Explorer: 10 名[*1]，Firefox: 1 名）．対照的に大学院生は，

[*1] 実験を実施した 2007 年 11 月時点では，タブブラウザ機能を追加した Internet Explorer 7 が公開されていたが，その利用率は 28.9%[41] にとどまっており，大半の参加者はタブブラウザ機能を持たない Internet Explorer 6 以下を使っていたと推測されるが，正確な常用ブラウザのバージョンに関しては確認していないため，不明である．

なんらかのタブブラウズ機能を持つブラウザを別途インストールして使用していると回答している（Sleipnir: 2 名，Firefox: 1 名，Opera: 1 名，不明: 1 名）．タブブラウジング機能は Web 探索を支援するために開発されたものであり，大学院生において積極的な利用行動が見られた点は，日常的に使っているブラウザの違いおよびその機能に対する習熟度に支えられていると見ることもできるが，このブラウザに対する習熟要因の影響の詳細な分析には至っていないため，今後の課題として残されている．

まとめると，大学院生は学部生に比べて，ページ閲覧の際のすばやい内容確認に加えて，ブックマークの積極的な付与を通じたより多くの有用ページのチェックやタブ機能による並列的な探索閲覧行動が特徴的であり，それらの行動を支えるために，検索バーやタブ機能といった比較的新しいブラウザ機能を多用するといった構図が見てとれる．

6.2 タスク種別が探索行動に与える影響

レポートタスクと旅行タスクによる探索行動の違いには，いくつかの種類の違いが見られた．

ひとつは，Lookzone 分析結果によれば，スニペットへの注視回数がレポートタスクのほうが多いことが示されており，この場合，レポートのための情報収集という性質から，ヒットした際の文書要約を通じて自らの情報要求に適合するサイトであるかどうかを読み取りながらの探索を行う傾向があったために，旅行タスクよりもスニペットへの注視が集まったと考えられる．

また，Lookzone などの注視分析およびクリックランクの分析結果によれば，サーチエンジンに関わる探索においてはレポートタスクのほうがスクロールバーへの注視も多いなど，より下位のランクへの閲覧が多くなる傾向があり，SERP からたどる際のクリックランクもより下位のものが多い傾向が示されている．Scanpath の長さもレポートタスクのほうが長く，この傾向を支持している．とくに，SERP ページの閲覧時間においてタスク間の差が無かったにもかかわらず，Scanpath での差があるという点からすると，SERP 内のランク箇所の読み取り方法においてタスク間に違いがあることが考えられる．

これらの結果はサーチエンジンの使い方においてレポート主題のような探索の場合であるか否かに応じて，閲覧ランクや読み行動の発生が左右されている可能性を示唆している．これはたとえば Broder が提唱したクエリ分類の軸 [19] のように，サーチエンジン利用のクエリが主題的な内容を含むのか，または単に特定のページへのナビゲーションを意図したようなクエリなのか，といった観点からも重要となる可能性を示している．

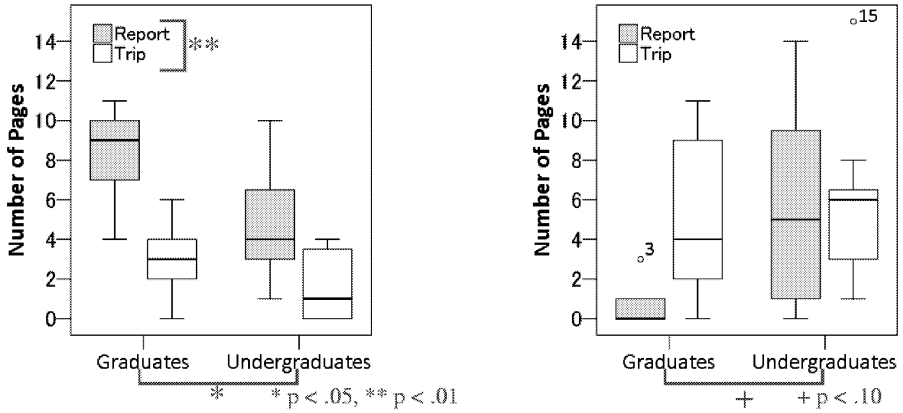
一方で，前節で述べたユーザ特性の違いと関連して，とくに学部生のレポートタスク遂行においては，ページ閲覧数が少なく，Link 行動も少なくなるなど，読み行動が多く発生している傾向が示されている．これは前節で述べたように大学院生と学部生のレポートタスクへの情報収集過程の違いとして見ることもできるが，タスク種別の違いとして見た場合には，レポートの執筆計画を立てる際に，テーマ選定から具体的な詳細化を行う過程において，ある程度の読み行動を必要とすることが多いということを示唆している．

また，レポートタスクよりも旅行タスクでの方が Link Depth 2 以上のページが多く閲覧されている．このことは，旅行タスクにおける閲覧パターンがよりサーチエンジンを離れた，個々のサイト内でのさらなる探索を必要とする傾向を示しており，つまりはサーチエンジンから個別サイトに移ったあとで，サイト内の情報を収集することが旅行タスクにおいては重要となっている可能性を示している．たとえば，旅行タスクがホテルや交通情報といった追加的な情報を求めるため，サイト内のナビゲーションをたどっていくタスクを内在していることに起因していると考えられる．

6.2.1 スポンサーリンク

もうひとつの側面として，SERP における Lookzone 分析結果からは，スポンサーリンクへの注視が旅行タスクのほうが多いことがわかる．

この観察の原因となりうる仮説として，1）旅行というタスクの性質から，スポンサーリンクに広告として掲載されるリンク先が重要な情報源となるため（スポンサーリンクの質），2）サーチエンジン上でスポンサーリンクが多く表示されるため（スポンサーリンクの量），の 2 つが考えられる．これらの仮説を裏付けるため，下記に述べる追加分析を行った．



(a) スポンサーリンク =0　　(b) スポンサーリンク ≧1

図 15　大学院生・学部生のタスクごとのスポンサーリンクが 0 個と 1 以上の SERP ページの閲覧回数

図 15 は，スポンサーリンクが 0 個の SERP ページと 1 以上の SERP ページの閲覧回数を示している．分散分析の結果，スポンサーリンクがまったく含まれない SERP ページでは，タスクの主効果に有意差があり（$F(1,14)=28.81$, $p < .01$），レポートタスクでは旅行タスクよりもスポンサーリンクの出現がない場合が多かった．また，グループの主効果にも有意差があり（$F(1,14)=5.76$, $p < .05$），大学院生は，学部生より多く，スポンサーリンクの含まれないページを閲覧していた．スポンサーリンクが 1 以上のページの閲覧回数では，グループ間に有意傾向があり，学部生グループのほうが大学院生よりもスポンサーリンクを含む SERP ページを閲覧していた（$F(1,14)=3.48$, $p < .10$）．

このことから，レポートタスクのほうがスポンサーリンクの出現と結びつかないクエリが多く，スポンサーリンクへの注視回数の旅行タスクとの差異を説明する大きな要因となっていることがわかった．

ただし，スポンサーリンクが出現したページ数としては，両タスク間に差はないため，最終的なスポンサーリンクへの注視行動で差がみられたことの原因は明らかとなっていないが，ユーザ自身の情報要求とスポンサーリンクとして表示されるタスク固有の情報資源とが合致した結果として旅行タスクでのスポンサーリンク注視を生起した可能性が考えられる．

7　おわりに

本研究では，ユーザの自然な情報探索行動を理解するために，ブラウザログ，画面キャプチャー映像，発話，眼球運動，インタビューなどの様々なデータを収集して，タスクや経験の違いが与える探索行動への影響を考察した．ユーザ実験から得られたデータにあわせ，COPATT によるタグ付け，Lookzone，Link Depth などの新たな分析手法を開発し，探索者のプロセスを量的・質的な両面から精緻に理解することを目指した．

タスク種別の影響としては，サーチエンジンにおける眼球運動データの分析から注視箇所の違いが明らかとなった．レポートタスクにおいてはスニペット部分への注視が旅行タスクに比べて多く，また旅行タスクにおいては，SERP 上におけるスポンサーリンクへの注視がより多く見られるなど，タスクにおいて必要とされる情報に応じて，着目する情報が異なることが示唆された．

ユーザ特性の側面では，Web 探索の経験と検索スキルの異なる 2 つのグループとして設定した，図書館情報学専攻の大学院生による探索プロセスと，一般の学部生の探索プロセス間の違いとしては，大学院生の方が全般にすばやく効率的なページ閲覧を行っており，学部生はそれに比べると比較的じっくりとページ内容を読み込みながらの判断を行っていた点が特筆できる．

また，大学院生グループの探索過程において，ブラウザのタブや検索バー機能を多用する利用法も，学部生グループとの顕著な差として観察された．

なお，探索中のプロトコル発話，探索後の事後インタビューといった探索者の内的状態を示すデータを用いた，より質的な分析については本論文の分析対象としてカバーできなかった．今後，上記で挙げたような分析手法に加えて，これらを用いた質的分析手法との組み合わせによる探索過程の解明を目指す予定である．

謝辞

本研究の一部は，科学研究費補助金基盤研究（B）（課題番号: 21300096）「検索ログ解析と認知的研究による利用者の探索的な情報検索行動の研究」，国立情報学研究所公募型共同研究「情報探索行動の認知モデルの構築とその応用に関する研究」の助成による．

参考文献

[1] Marchionini, Gary: “Exploratory search: from finding to understanding”, *Communications of the ACM*, Vol. 49, No. 4, pp. 41–46, 2006.

[2] Spink, Amanda; Jansen, Bernard. J.: “Web Search: Public Searching of the Web”. Kluwer Academic Publishers, Dordrecht, the Netherlands, 199p., 2004.

[3] Terai, Hitoshi; Saito, Hitomi; Egusa, Yuka; Takaku, Masao; Miwa, Makiko; Kando, Noriko: “Differences between informational and transactional tasks in information seeking on the web”, *Proceedings of IIiX 2008*, pp. 152–159. ACM, 2008.

[4] Egusa, Yuka; Takaku, Masao; Terai, Hitoshi; Saito, Hitomi; Kando, Noriko; Miwa, Makiko: “Visualization of User Eye Movements for Search Result Pages”, *Proceedings of EVIA 2008 (NTCIR-7 Pre-Meeting Workshop)*, pp. 42–46, 2008.

[5] 齋藤ひとみ; 江草由佳; 高久雅生; 寺井仁; 三輪眞木子; 神門典子:「Web 情報探索行動の分析: 課題の志向性と経験の違いによる影響についての予備的検討」, 電子情報通信学会研究報告「Web インテリジェンスとインタラクション」研究会 (IEICE SIG-WI2) 第 13 回研究会, pp. 37–42, 2008.

[6] 高久雅生; 寺井仁; 江草由佳; 齋藤ひとみ; 三輪眞木子; 神門典子:「Web 情報探索における視線データの予備的分析」, 情報知識学会誌, Vol. 18, No. 2, pp. 181–188, 2008.

[7] 高久雅生; 江草由佳; 寺井仁; 齋藤ひとみ; 三輪眞木子; 神門典子:「サーチエンジン検索結果ページにおける視線情報の分析」, 情報知識学会誌, Vol. 19, No. 2, pp. 224–235, 2009.

[8] 江草由佳; 高久雅生; 寺井仁; 齋藤ひとみ; 三輪眞木子; 神門典子:「LinkDepth: Web 情報探索行動の閲覧パターンの分析」, 情報処理学会研究報告, Vol. 2009-FI-95, No. 20, pp. 1–7, 2009.

[9] 三輪眞木子; 江草由佳; 齋藤ひとみ; 高久雅生; 寺井仁; 神門典子:「Web 上の exploratory search の特徴: 発話プロトコルと事後インタビュー分析結果より」, 情報処理学会研究報告, 第 2009-FI-96 巻, pp. 1–8, 2009.

[10] Miwa, Makiko; Kando, Noriko: “Naïve Ontology for Concepts of Time and Space for Searching and Learning”, *Information Research*, Vol. 12, No. 2, 2007. URL: `http://informationr.net/ir/12-2/paper296.html`.

[11] Miwa, Makiko; Kando, Noriko: “Methodology for Capturing Exploratory Search Processes”, *Proceeding of the ACM SIGCHI 2007 Workshop on Exploratory Search and HCI : Designing and Evaluating Interfaces to Support Exploratory Search Interaction (ESI2007)*, pp. 76–80, San Jose, USA, 2007.

[12] Miwa, Makiko; Kando, Noriko: “Role of Naïve Ontology in Search and Learn Processes for Domain Novices”, *Proceedings of the 9th International Conference on Asian Digital Libraries, ICADL 2006*, pp. 380–389, Kyoto, Japan, 2006.

[13] Kellar, Melanie; Watters, Carolyn; Shepherd, Michael: "A field study characterizing Web-based information-seeking tasks", *Journal of the American Society for Information Science and Technology*, Vol. 58, No. 7, pp. 999–1018, 2007.

[14] Saracevic, Tefko; Kantor, Paul: "A study of information seeking and retrieving. II. Users, questions, and effectiveness", *Journal of the American Society for Information Science*, Vol. 39, No. 3, pp. 177–196, 1988.

[15] Kim, Kyung-Sun; Allen, Bryce: "Cognitive and task influences on Web searching behavior", *Journal of the American Society for Information Science and Technology*, Vol. 53, No. 2, pp. 109–119, 2002.

[16] Thatcher, Andrew: "Web search strategies: The influence of Web experience and task type", *Information Processing & Management*, Vol. 44, No. 3, pp. 1308–1329, 2008.

[17] Marchionini, Gary; Shneiderman, Ben: "Finding Facts vs. Browsing Knowledge in Hypertext Systems", *Computer*, Vol. 21, No. 1, pp. 70–80, 1988.

[18] Byström, Katriina; Järvelin, Kalervo: "Task complexity affects information seeking and use", *Information Processing & Management*, Vol. 31, No. 2, pp. 191–213, 1995.

[19] Broder, Andrei: "A Taxonomy of Web Search", *SIGIR Forum*, Vol. 36, No. 2, pp. 3–10, 2002.

[20] Hembrooke, Helene A.; Granka, Laura A.; Gay, Geraldine K.; Liddy, Elizabeth D.: "The effects of expertise and feedback on search term selection and subsequent learning: Research Articles", *Journal of the American Society for Information Science and Technology*, Vol. 56, No. 8, pp. 861–871, 2005.

[21] Moore, Joi L.; Erdelez, Sanda; He, Wu: "The Search Experience Variable in Information Behavior Research", *Journal of the American Society for Information Science and Technology*, Vol. 58, No. 10, pp. 1529–1546, 2007.

[22] Hölscher, Christoph; Strube, Gerhard: "Web search behavior of Internet experts and newbies", *Proceedings of the 9th international World Wide Web conference*, pp. 337–346, 2000.

[23] Lazonder, Ard W.; Biemans, Harm J. A.; Wopereis, Iwan G. J. H.: "Differences between novice and experienced users in searching information on the World Wide Web", *Journal of the American Society for Information Science and Technology*, Vol. 51, No. 6, pp. 576–581, 2000.

[24] Terai, Hitoshi; Miwa, Kazuhisa: "Sudden and Gradual Processes of Insight Problem Solving: Investigation by Combination of Experiments and Simulations", *Proceedings of 28th Annual Meeting of the Cognitive Science Society*, pp. 834–839, 2006.

[25] Jacob, Robert. J. K.; Karn, Keith. S.: "Eye tracking in human–computer interaction and usability research: ready to deliver the promises (section commentary)". Hyona, Jukka; Radach, Ralph; Deubel, Heiner, editors, *The mind's eye: cognitive and applied aspects of eye movement research*, pp. 573–605. Elsevier Science, Amsterdam, 2003.

[26] Anderson, John R.; Gluck, Kevin A.: "What role do cognitive architectures play in intelligent tutoring systems?". Carver, S. M.; Klahr, D., editors, *Cognition and Instruction: Twenty-five Years of Progress*, pp. 227–262. Lawrence Erlbaum Associates, 2001.

[27] Kelly, Diane: "Implicit Feedback: Using Behavior to Infer Relevance". Spink, Amanda; Cole, Charles, editors, *New Directions in Cognitive Information Retrieval*, pp. 169–186. Springer, Dordrecht, 2005.

[28] Nielsen, Jakob: "Eyetracking Research

into Web Usability". URL: http://www.useit.com/eyetracking/ (2009年4月10日参照).

[29] Lorigo, Lori; Haridasan, Maya; Brynjarsdottir, Hronn; Xia, Ling; Joachims, Thorsten; Gay, Geri; Granka, Laura; Pellacini, Fabio; Pan, Bing: "Eye tracking and online search: Lessons learned and challenges ahead", *Journal of the American Society for Information Science and Technology*, Vol. 59, No. 7, pp. 1041–1052, 2008.

[30] Granka, Laura A.; Joachims, Thorsten; Gay, Geri: "Eye-tracking analysis of user behavior in WWW search", *SIGIR '04: Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 478–479. ACM, 2004.

[31] Guan, Zhiwei; Cutrell, Edward: "An eye tracking study of the effect of target rank on web search", *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 417–420. ACM, 2007.

[32] Pan, Bing; Hembrooke, Helene; Joachims, Thorsten; Lorigo, Lori; Gay, Geri; Granka, Laura: "In Google we trust: Users' decisions on rank, position and relevancy", *Journal of Computer-Mediated Communication*, Vol. 12, No. 3, pp. 801–823, 2007.

[33] Lorigo, Lori; Pan, Bing; Hembrooke, Helene; Joachims, Thorsten; Granka, Laura; Gay, Geri: "The influence of task and gender on search and evaluation behavior using Google", *Information Processing & Management*, Vol. 42, No. 4, pp. 1123–1132, 2006.

[34] Saracevic, Tefko: "Relevance reconsidered", *Proceedings of the Second Conference on Conceptions of Library and Information Science, Information science: Integration in perspectives*, pp. 201–218, Copenhagen, Denmark, 1996.

[35] Saracevic, Tefko: "Relevance: A Review of the Literature and a Framework for Thinking on the Notion in Information Science. Part II: Nature and Manifestations of Relevance", *Journal of the American Society for Information Science and Technology*, Vol. 58, No. 13, pp. 1915–1933, 2007.

[36] 齋藤ひとみ; 三輪和久: 「問題解決活動としての WWW 情報探索 : 科学的発見の枠組みに基づく検討」, 認知科学, Vol. 10, No. 2, pp. 258–275, 2003.

[37] Joachims, Thorsten: "Optimizing search engines using clickthrough data", *Proceedings of KDD 2002*, pp. 133–142. ACM, 2002.

[38] Dupret, Georges E.; Piwowarski, Benjamin: "A user browsing model to predict search engine click data from past observations.", *SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 331–338. ACM, 2008.

[39] Baeza-Yates, Ricardo; Calderón-Benavides, Liliana; González-Caro, Cristina: "The Intention Behind Web Queries", *Proceedings of SPIRE 2006*, pp. 98–109, 2006.

[40] 斎田真也: 「読みと眼球運動」. 『眼球運動の実験心理学』. 第 8 章, pp. 167–197. 名古屋大学出版会, 名古屋, 1993.

[41] 財団法人インターネット協会監修: 「インターネット白書 2008」. インプレス R&D, 東京, 368p., 2008.

(2009年11月26日受付)
(2010年4月27日採択)
(2010年5月24日オンライン公開)

論文

Poker-Maker モデル：ユーザの検索意図を反映するキーワードマップと情報収集エージェントの連携による探索的情報検索

Poker–Maker model: Exploratory search by collaboration between Keyword Map reflecting user's search intention and information gathering agent

梶並知記[1*], 高間康史[1]
Tomoki KAJINAMI[1*], Yasufumi TAKAMA[1]

1 首都大学東京
Tokyo Metropolitan University
〒191-0065 東京都日野市旭が丘6-6
E-mail: kajinami@krectmt3.sd.tmu.ac.jp

本稿では，探索的情報検索を伴う多様なタスクを支援するために，対話的な汎用的情報分析ツールであるキーワードマップとエージェントアプローチを統合したシステムアーキテクチャを提案する．エージェントアプローチに基づく既存の情報収集システムは，柔軟なシステム設計が可能でユーザの多様な要求に応えることができるが，明確な検索要求を前提にしているため，情報分析と探索的情報検索を通して要求を精緻化し意思決定を行う作業への対応は困難である．本稿では，エージェントアプローチの利点を活かしつつ情報分析作業までを包括的に支援可能とする．汎用的な情報分析を可能とするキーワードマップを共通のインタフェースとして採用し，ユーザの情報分析操作からユーザの検索意図を抽出し，これに応じて情報収集・可視化を行うPoker-Makerモデルを提案する．提案モデルを用いた3種類の異なる意思決定タスクに対する既存システムの理解により，モデルの妥当性を示す．

This paper proposes the integration of Keyword Map (an interactive information visualization tool) and an agent-based system architecture for supporting exploratory searches. Keyword Map can be used for analyzing data from various perspectives. A user can suitably arrange keywords using interaction functions of the system in order to reflect his/her data analysis requirements. Typically, an information gathering agent searches for information according to the user's specific requirement and generates data for presenting search results on the interface. Although agent-based information search systems have been developed to be flexible, they cannot support the user's exploratory search process because such a process cannot explicitly reflect the information requirement. By employing Keyword Map as the interface between a user and the information gathering agent, the proposed approach aims to connect the process required for the user's data analysis with the agent's retrieval process. The usa-

bility of the proposed approach has been demonstrated by applying it for existing search systems for three different types of tasks.

キーワード: インタラクティブ情報検索，ユーザ意図，情報可視化，情報収集エージェント，キーワードマップ
Interactive information retrieval, User's intention, Information visualization, Information gathering agent, Keyword Map

1 はじめに

インターネットインフラの整備に伴い，我々は容易に多量の情報へのアクセスが可能となっているが，同時に以下の2 つの難題に直面している．

問題1：必要としている情報の場所がわからず辿りつくのが困難

問題2：多量の情報を比較検討する分析作業に対する負荷が増大

問題1を解決するために，ユーザが簡単な命令を与えれば，全自動もしくは半自動的にユーザの要求に応じた情報を収集してくる情報収集エージェントに関する研究が行われている[1][2]．エージェントアプローチは，役割の異なる複数のエージェントを連携させ，タスクに応じてエージェントの構成を変更することで，複数の情報源を利用する多様なタスクに対応できる利点がある．エージェント同士の相互作用や，情報収集の複雑な過程などはユーザから隠されるため，ユーザは，インタフェースのみを利用して自身の要求をエージェントへ伝達できる利点もある．

問題2を解決するためには，情報可視化[3]が有効であることが知られており，ユーザが能動的にデータを眺めることを可能にするインタラクティブな情報可視化システムが提案されている[4][5][6]．キーワードマップはそれらのシステムの一種であり，2D平面に可視化されたキーワード空間をユーザが編集することにより，データ分析に関する意図をキーワード空間へ反映可能である[7][8]．キーワードマップは，データ空間全体（キーワード空間全体）の，一部分を可視化する． 本稿の目的は，既存エージェントアプローチがもつ利点を活かしつつ，さらに意思決定における情報分析作業までを包括的に支援可能となるように拡張することである．

意思決定における情報収集は，ユーザの検索要求が不明確であり，探索的情報検索と検索結果の分析を通して，要求を精緻化していくといった特色がある．エージェントアプローチでは，ユーザの要求が比較的明確に定まっている状況での利用を想定しているため，不明確な要求が情報分析作業を通じて精緻化されていくプロセスを考慮していない．また，ユーザは，提示された情報を別途分析し，改めて明確な要求をエージェントに送ることになり，分析作業と検索作業を分けて行わなくてはならない．したがって，既存の情報収集エージェントをそのまま用いることはできない．さらに，情報分析作業から検索要求を抽出する必要がある．

キーワードマップは汎用的な情報分析インタフェースであるが，ユーザの意図に応じた情報検索やデータセットの作成（タスクに依存している部分）を行う外部システムとどのように連携させれば検索システム全体を設計できるのか，その設計の指針が示されていない．検索システム全体の設計を示すことで，システム開発の際に，多様なデータ，タスクを対象とした意思決定支援システムを同一のフレームワークで構築可能になる．

キーワードマップはタスクに応じて複数のエ

ージェントと連携することができるため，エージェントアプローチの利点が活かされる．また，タスクが異なっても共通のインタフェースを用いて情報収集エージェントと連携し問題解決にあたることができ，ユーザは，複数の異なるインタフェース操作に習熟する必要がなくなる．

キーワードマップは，キーワードマップ上でのユーザの操作（データ分析操作）ログを外部に出力でき，データセットを外部から受け取る．外部システムと通信してやり取りされるものは，データ分析操作ログとデータセットの2つのみである．したがって，キーワードマップとエージェントアプローチを組み合わせるために解決すべき点は以下の2点となる．

・データ分析操作からの検索意図の抽出
・検索意図に応じた可視化データの生成

ユーザがデータ分析中に自身の分析意図と検索意図を意識して区別する必要がなくなるためには，情報収集エージェントに依存しないデータ分析操作の枠組みが必要である．また，情報収集エージェントが的確に検索意図を抽出しそれに基づいて検索するためには，ユーザのデータ分析操作から直接これらの作業を行う枠組みと，ユーザによるデータ分析操作に対応可能なデータセットを構築する枠組みとが必要である．そのためには，ユーザと情報収集エージェントの操作対象を共通にして，操作法に統一性を持たせる必要がある．

情報収集エージェントへユーザの検索意図を正しく伝達するために，キーワードマップ上での操作からユーザの検索意図を抽出する手法について検討する．ユーザの，検索意図を含んだデータ分析操作（キーワードマップ上で行う）を，意思決定のためにデータを多角的な視点から分析するOLAP[9][10]に対応づけ，3次元データキューブでキーワードマップ用データセットを表すKMキューブを提案する．本稿では，

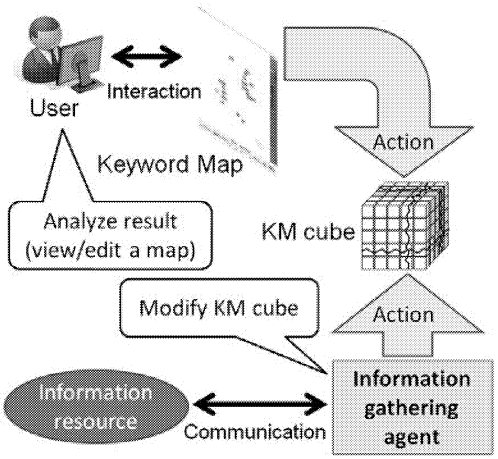


図 1　提案モデルの概要

ユーザによるデータ分析操作から対応する検索意図を抽出して情報収集エージェントへ伝達し，情報収集エージェントが，ユーザの検索意図に応じてKMキューブを再構成していく枠組みを，Poker-Maker（Point KEywords and Relevance balance setting - Modify Active Keyword map cube Expressing Relevance data）モデルとして提案する．提案モデルの概要を，図1に示す．

KMキューブはキーワードマップと情報収集エージェントの間におかれ，ユーザと情報収集エージェントがKM キューブを操作する．提案モデルは，インタフェースを統一し，分析作業と検索作業の区別を無くし，かつ汎用的なタスクに対応可能な，意思決定のための探索的情報検索支援システムの設計に役立つ．また，既存システムの理解にも役立つ．提案モデルの基本コンセプトは，以下の3点である．（1）キーワードマップで可視化されているのは，データ空間全体の一部であるため，マップ上での（局所的な）データ分析操作は，現在のマップに含まれていないが関連するデータを検索したいという意図も含んでいると考える．（2）ユーザは，キーワードマップを用いKMキューブに対してOLAPと同様の分析操作を行うことで，検索意

図を表現する．(3)情報収集エージェントは，ユーザの検索意図に応じて，KMキューブに対してOLAP操作を行いKMキューブを再構築する．

本稿で対象とする探索的情報検索は，ユーザが情報空間の一部である検索結果を局所的にデータ分析した結果，クエリとなるキーワードとキーワードに関する視点を変更して再検索を繰り返し行うことで，ユーザが自身の要求を精緻化しつつ検索を行うことである．そのためコンセプト (1) が設定可能になる．また，OLAPは意思決定のために多角的な視点からデータ分析を行う手法としてよく知られているため，OLAP同様の操作でKMキューブを操作するコンセプト (2) (3) を設定する．

情報可視化技術を用いたインタフェースと外部システムを連携して検索を支援する関連研究として，INSYDER[11]が提案しているが，これは分析作業と検索作業(クエリの生成作業)で異なる可視化表現を用いており，分析と検索作業を同一の可視化表現で一体化する本稿の提案モデルと異なる．また，インタフェースを一般的なWeb ブラウザに統一(ブラウザに組み込むプラグインとして実装)し，ユーザの検索意図を考慮した検索支援システムも提案されている[12]．このシステムでは，ユーザが指定した単語の意味に基づき，いくつかの検索方法をユーザに提示する．その中から，ユーザは，検索方法を別途選択することになる．ユーザの分析操作から検索意図を抽出し，ユーザが分析と検索を意識して区別する必要のない本稿の提案モデルとは異なる．

本稿の構成は以下のとおりである．2節では，本稿で想定する情報収集エージェントに関連する研究について述べる．3節では，キーワードマップを用いた対話的データ分析方法について述べる．4節では，OLAP で使われる多次元データキューブのサブセットとして，キーワードマップ用データセットを表すKMキューブを提案する．5節では，ユーザによるデータ分析操作から対応する検索意図を抽出して情報収集エージェントへ伝達し，情報収集エージェントが，ユーザの検索意図に応じてKMキューブを再構成していく枠組み，Poker-Makerモデルを提案する．6節で，3種類の異なる意思決定タスクに対する既存システムを，提案モデルによって理解することにより，モデルの妥当性を示す．

2 情報収集エージェントに関する研究

本稿で想定する情報収集エージェントは，ユーザの検索意図を含んだシードとなる情報が与えられた場合に，それに応じて情報源から情報を抽出し，結果の提示手法に合わせたデータセットやデータベースを作成する役割を持つものである．これは，一般的な情報収集エージェントの役割である．既存研究においても，ユーザが検索を通してある概念を理解する作業を支援するために，Web ページや概念を表す単語の集合をシードにして関連ページを収集し，より概念の理解に役立つページを判定しデータベースへ格納するエージェントが提案されている[13][14]．

また，ユーザの要求を満たすアイテムを推薦するため，要求を満たすWeb 上の関連情報を収集し，アイテムデータベースに登録したり，黒板オブジェクトとして階層構造を持つ黒板に格納したりするエージェントも提案されている[15][16]．

これらの研究は，広い意味で意思決定タスクの支援であるが，ユーザの要求が比較的定まっている場合を想定していたり，ユーザによる検索結果の簡単な評価から，ユーザの嗜好を推定して検索を行ったりしている．ユー

ザの情報分析作業そのもの支援や，ユーザが分析を通してユーザの要求を精緻化していく過程はあまり想定されていない．

3 キーワードマップによる対話的なデータ分析

キーワードマップは，汎用的なデータ分析・意思決定支援を目的とした，対話的な情報可視化システムである[17]．

キーワードマップは，可視化対象となるオブジェクトに付与されたラベルであるキーワードの集合S_wと，それらの間の関連度からなるデータセットを読み込み，バネモデル[18]を用いて2D平面上にキーワードを自動配置する．関連度はキーワード間の関連性を表す複数の属性（関連属性r^k）に重みを付けて線形結合したものである．バネの長さを非関連度に比例して定義することで，関連度の高いキーワード程近くに配置される．

キーワードマップは，ユーザによる任意タイミングでのキーワード配置への介入を許す，多数のインタラクション機能を備えている．これらを用いることで，対話的なデータ分析が可能である．特徴的なインタラクション機能は，以下に示す「注目/削除キーワード選択機能」「意図強調機能」「関連バランス制御機能」の3つの機能に大別される．また，ユーザがインタラクション機能を用いてキーワード配置に介入することを，「マップ編集」と呼ぶ．

・注目/削除キーワード選択機能

ユーザが選択したキーワードを，注目キーワードw_f，もしくは削除キーワードw_dとして，バネモデルに基づく配置を無視して任意の場所へ移動・固定する機能である．注目キーワードw_fは，次に述べる意図強調機能を使用する際の対象キーワードとなる．また，注目キーワードに繋がるリンクが実線で表示される．なお，削除キーワードw_dに関しては，それに繋がるバネも一時的に全て削除され，意図強調機能の対象にはならない．

・意図強調機能

ユーザのプリミティブなデータ分析意図[7]（複数のキーワードを同一話題として扱いたい，異なる話題として扱いたい，重要度に差をつけて扱いたい）に応じて，注目キーワードw_fを中心とするクラスタを生成したり，分離したり，整列させたりして，ユーザ意図を強調したキーワード配置の作成を支援する機能である．

・関連バランス制御機能

関連属性r^kの重要度を，関連属性の重みg^kを制御することで自由に変更し，キーワード配置に反映する機能である．以後，本稿では，ユーザが任意の関連属性の重みを制御することを「関連バランス制御」と呼び，関連属性の重みg^kを要素とするベクトル$\boldsymbol{G_r}$を「関連バランス設定(Relevance balance setting)」と呼ぶ．関連バランス制御には，意思決定方略[19]に応じたデータ分析の観点から，以下のような3種類のパターン（極大化，極小化，傾斜化）が使われる[17]．

極大化：任意の関連属性r^kの重みg^kを，他の属性の重みが無視できるほど大きくすることである．これにより，r^k以外を無視した視点でデータを眺めることができる．

極小化：任意の関連属性r^kの重みg^kを，他の属性と比較して無視できるほど小さくすることである．これにより，r^kを無視した視点でデータを眺めることができる．

傾斜化：極大化・極小化以外の状態であり，複数の関連属性の重要度に差をつけてデータを眺めることができる．

なお，システム上の扱いでは，極大化・極

小化も，G_rの任意の要素を，他の要素と比較して十分に大きくしたり小さくしたりして表現できる．

図2は，4つの関連属性を含むデータを読み込み，システムによってキーワードが自動配置された後，ユーザがマップ編集を行ったキーワードマップの例である．マップA（Map A），B（Map B）は共に，意図強調機能を使用し，2つの注目キーワード（w_{f1}，w_{f2}）に関連するキーワードを集めて2つのクラスタ（実線，破線の円）を作成している．マップAでは，3つの関連属性を傾斜化，マップBでは1つの関連属性を極大化している．2つのマップで，w_{f1}，w_{f2}は変化していないが，クラスタ内の他のキーワードが両マップで異なっている．また，マップBでは，両方のクラスタに関連するキーワードクラスタ（実線の短形）が生成されている．図2より，関連バランス制御機能を用いることで，異なった視点からキーワード間の関連性を眺めることが可能なことがわかる．

キーワード配置は，ユーザの操作に応じてリアルタイムにアニメーションを伴い動的に変化する．キーワードの移動先や関連バランス設定を決定してからキーワードの再配置が一括して行われるわけではなく，マウスボタンをクリックしたりドラッグしたり，マウスホイールを回転させる操作に応じて，スムーズに配置が変化する．これらのインタラクション機能により，自身の曖昧な意図を強調・外在化することや，様々な視点からデータを分析することができるため，キーワードマップは，データ分析意図の異なる様々な意思決定タスクに有効である[8]．

キーワードマップはデータの可視化と，ユーザとのインタラクションを担当し，データ

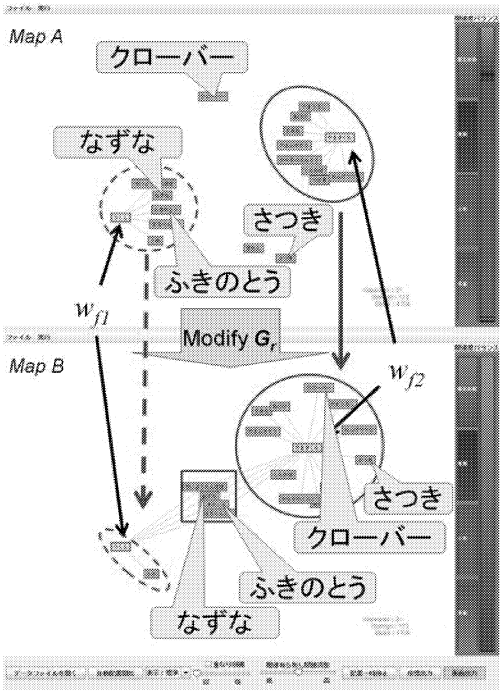


図 2 関連バランス設定に応じたキーワード配置の変化

セットは外部システムで作成できる．そのため，タスクに応じて外部システムを切り替えることで，インタフェースを同一のキーワードマップとしたまま，様々なタスクに対応できる．この特徴から，キーワードマップは，タスクに応じてエージェントの組み合わせを変えるエージェントアプローチとの親和性があるといえる．

4 KMキューブ

4.1 KMキューブの位置づけと構築

KMキューブは，ユーザの対話的なデータ分析作業を支援するインタフェースであるキーワードマップと，情報源からユーザの検索意図に応じて情報を収集する情報収集エージェントを仲介するものである．ユーザは，マップ編集を通してKMキューブを操作することでデータを分析する．この分析

には，新たに情報を検索したいという意図も含まれる．情報収集エージェントはユーザの KM キューブ操作からユーザの検索意図を抽出し，これに応じて情報を再検索後，新たな KM キューブを構築する．キーワードマップと情報収集エージェントの間に KM キューブを挟むことで，データ分析を伴った探索的情報検索作業における，キーワードマップと情報収集エージェントの役割分担を明確にすることが可能である．さらに，両者のアクションを，KM キューブに対するものとして統一的に取り扱うことができる．これにより，ユーザはデータ分析と検索要求を意識して区別する必要がなくなり既存アプローチの問題点を解決できるほか，情報収集エージェントはユーザのデータ分析操作から直接，検索意図を抽出できるといった利点がある．また，KM キューブを多次元データキューブとしてモデル化することで，KM キューブに対するアクションを OLAP 操作として整理できる利点もある．

キーワードマップは，情報空間の一部を可視化し，情報空間を任意のキーワード対（i, j）とそれらの間の関連属性r^k_{ij}で扱う．そのため，KMキューブは，3次元のデータキューブとして情報空間からダイシングすることで構築する．KMキューブは，3つの軸D_{w1}, D_{w2}, D_r をもち，D_{w1}, D_{w2} はキーワード集合S_w に，D_r は関連属性に対応し，データ値は対応するキーワード間の関連属性値である．3次元で固定される点と，2つの軸（D_{w1}, D_{w2}）の要素が必ず同じである点が，OLAPで用いられる多次元データキューブと異なる．

4.2 ユーザによるマップ編集を通したKMキューブへのアクションと検索意図

3節で紹介したキーワードマップ上の編集

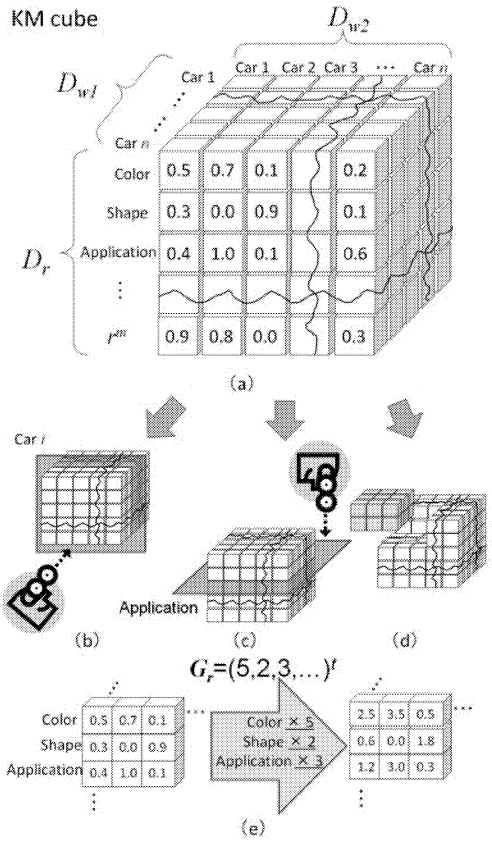


図 3　KM キューブ

操作（インタラクション機能）は，このKMキューブへのOLAP操作に対応づけられる．図3 (a) は，新車購入を考えているユーザが多種類の車 (Car 1～Car n) を分類・比較して，購入する車を選択するタスクを想定した場合の，KMキューブの例である．関連属性には，色（Color）や形状（Shape），主な用途（Application）などがあり，関連属性値がデータ値として格納されている．ユーザが，特定のキーワードw_fとその他のキーワードの関連性を分析することは，注目キーワード選択機能を用いてw_fを選択し，図3 (b) のようにKMキューブを任意の要素iでスライス (Slice) しデータを眺めることに対応する．ここで，「データを眺める」とは，切り出さ

れたデータ行列に含まれる関連属性値を強調したキーワード配置をマップ上で行うことを意味する．同様に，特定の関連属性r^kを極大化してキーワードの関連性を分析することは，関連バランス制御機能を用いてr^kを極大化し，図3（c）のようにKMキューブを「用途（Application）」でスライスし，データを眺めることに対応する．また，複数の注目キーワード$\{w_{fi}\}$や関連属性$\{r^k\}$を選択する場合や，削除キーワード選択機能を用いる場合，関連属性を極小化する場合は全て，図3（d）のようにKMキューブをダイシング（Dice）することに対応する．関連属性の傾斜化は，関連バランス設定$\boldsymbol{G}_r$を用いて該当の関連属性値に重み付けを行う演算（Calculation）に対応する．図3（e）では，色属性の重みを他の属性より重くした関連バランス設定$\boldsymbol{G}_r$による演算の例である．なお，OLAPの場合と同じく，これらの操作は組み合わせて用いることができる．

キーワードマップで可視化されているのは情報空間全体の一部分である．そのため，マップ上で行われる分析を局所的な分析とみなすことができる．また，検索により新たな分析対象空間を求めることを大局的な分析とみなすことができる．しかし，ユーザの観点から見ると，マップ編集中に，分析意図，検索意図とは必ずしも分離されていない．したがって，ユーザによるマップ編集操作には，上で述べたような，現在のマップをどのような観点から眺めたいかといった，局所的な分析に関する意図だけでなく，大局的な分析に対応する検索意図も込められていると考える．

本稿では，以下に示す5つの分析・検索に関する意図を，タスクに依存しないプリミティブなものとして想定する．

意図K1：注目キーワードw_fと関連のあるキーワードを見たい．
意図K2：あるキーワードw_dを無視したい．
意図R1：ある関連属性r^kに絞って関連のあるキーワードを見たい．（極大化）
意図R2：ある関連属性r^kを無視して関連のあるキーワードを見たい．（極小化）
意図R3：複数の関連属性を考慮して関連のあるキーワードを見たい．（傾斜化）

これらは，キーワードに着目した意図（K1，K2）と関連属性に着目した意図（R1～R3）に大別できる．K1とK2は，探索的情報検索中に変更されるクエリのキーワード選択に対応し，R1～R3は，キーワードに関してどのような視点を重視するのかといった意思決定方略に対応する．なお，K1，K2においては複数のキーワードを指定することも含む．これらの意図に対し，ユーザがキーワードマップ上で行う操作は，K1，R1はスライシング（複数指定の場合はダイシング），K2，R2 はダイシング，R3は演算に対応する．これらの操作では，キーワード集合S_wそのものは変化しない．

なお，5つの意図に関する，局所的な意図と大局的な意図の関係は，意図の対象が，現在可視化されている情報空間の一部分から，異なる情報空間へ変化することに対応することとなる．すなわち，5つの意図の詳細説明それぞれの最後が，「見たい」から「検索したい」と変化することになる．

また，キーワードマップは多角的な視点からのデータ分析を可能とするが，関連バランス設定$\boldsymbol{G}_r$はマップ全体へ影響する．そのため，複数のキーワードを含むKMキューブのダイシングは可能だが，キーワードAに関しては関連属性

α，キーワードBに関しては関連属性βで極大化するような，同時に複数の関連バランス設定を行う演算はできない．

4.3 情報収集エージェントによるKMキューブへのアクション

4.2節の操作は，構築されたKMキューブに対してユーザが行うのに対し，情報収集エージェントが行う，ユーザの検索意図に対応した操作は，データ空間全体に対応した多次元データキューブから，ユーザの要求に関連した部分空間をKMキューブとしてダイシングすることに相当する．ユーザの意図に対し，ユーザとエージェントがそれぞれ別の操作を行うことで，検索意図を実現する．4.2節で述べた5つの意図それぞれについて，エージェントが関与するのは検索に関するものである．

ユーザが意図K1，K2を持つ場合，情報収集エージェントは，ユーザが注目したキーワードに対し，指定された関連属性のいずれかにおいて関連するキーワードを収集し，S_wとする．情報収集エージェントの操作は，D_{w1}，D_{w2}を構成する要素（キーワード）S_wを選択（収集）するダイシングである．意図R1～R3はD_rを構成する要素についてのダイシングに対応するが，通常は局所的分析で扱えばよく，情報収集の段階で属性を取捨選択する必要はないと考える．しかし，6.3節で紹介する事例のように，分析対象とするキーワード数を絞り込みたい場合などには，これらの意図に基づく操作は効果的と考える．また，6.2節で紹介する事例のように，関連属性数が多い場合には，KMキューブ再構築時に，ロールアップに相当する操作を考慮することが望ましい．

以上のように，ユーザのデータ分析意図を，局所的な意図と，それに対応した大局的な分析に対応する検索意図に分割して扱うことにより，ユーザによるデータ分析作業とエージェントによる情報収集作業を切り分けることができるだけでなく，ユーザによるマップ編集操作から検索意図を自動的に抽出できるといった利点もある．

4.4 KMキューブからのキーワードマップ用データセットの作成

キーワードマップ用のデータセット（可視化対象となるオブジェクトのキーワード集合S_wと，それらの間の関連度からなる）は，4.3節によって情報収集エージェントがダイシングしたKMキューブから次の手順で作成される．（1）S_w中のキーワードの全ての対をバネで結ぶ．（2）（1）のバネに，キーワードi，j間の関連属性値$r^k_{i,j}$を関連付け，データセットとする．可視化の際には，$r^k_{i,j}$に重みを付けて線形結合した$R_{i,j}$を，[0, 1]の範囲で正規化して関連度とし，バネの長さは非関連度（1− $R_{i,j}$）に比例したものとする[17]．

キーワード集合 S_w に含めるキーワードの決定や，関連属性値の決定は，アプリケーションごとに行うことができる．

5 Poker-Makerモデルの提案

4.2節より，ユーザの様々な検索意図を表現するマップ編集結果は，キーワードへの指示と，関連属性への指示の2つで表せる．したがって，キーワードマップから情報収集エージェントに対し，ユーザの選択したキーワード集合$\{w_{fi}\}$と$\{w_{di}\}$，関連バランス設定$\boldsymbol{G_r}$を送るだけで，ユーザの意図に応じた再検索と新たなKMキューブの構築による，再検索結果の可視化が可能になる．ユーザによるデータ分析操作から対応する検索意図を抽出し

て情報収集エージェントへ伝達し，情報収集エージェントが，ユーザの検索意図に応じてKMキューブを再構成していく枠組みを，Poker- Maker（POint KEywords and Relevance balance setting - Modify Active Keyword map cube Expressing Relevance data）モデルとして提案する．

提案モデルは，ユーザ（User）がキーワードマップ（Keyword Map）を通しKMキューブ（KM cube）へアクションを行い，情報収集エージェント（Information gathering agent）も，KMキューブへアクションを行う．4.2節で述べたユーザの意図に応じた，ユーザと情報収集エージェントそれぞれのKMキューブへのアクションは，表1にまとめた．表中の括弧内のアクションは，場合によって行うこともあるものである．また，関連属性への指示を，同時に複数行うことはできない．

図4はPoker- Makerモデルの概要と，データの流れを示したものである．実際にシステムを構成した場合のデータの流れは，図4中の破線矢印となる．

このモデルは，以下のステップで探索的情報検索を支援する．

（1） ユーザによる，情報収集エージェントへの初期クエリの入力

（2） 情報収集エージェントによる処理
- 検索結果からKMキューブの構築
- キーワードマップ用データセットの作成

（3） キーワードマップによるデータセットの可視化

（4） ユーザによる，データ分析（マップ編集）

（5） キーワードマップによる，マップ編集ログの出力

（6） 情報収集エージェントによる処理
- マップ編集ログからユーザの検索意図の抽出
- ユーザの検索意図に応じた検索

（7） ステップ2へ

表1 ユーザの意図に応じたユーザとエージェントのKMキューブへのアクション

User's search intention	User's action	Agent's action
Intention K1	Slice (Dice)	Dice
Intention K2	Dice	Dice
Intention R1	Slice (Dice)	Dice (Drill-down)
Intention R2	Dice	Dice (Roll-up)
Intention R3	Calculation	Dice

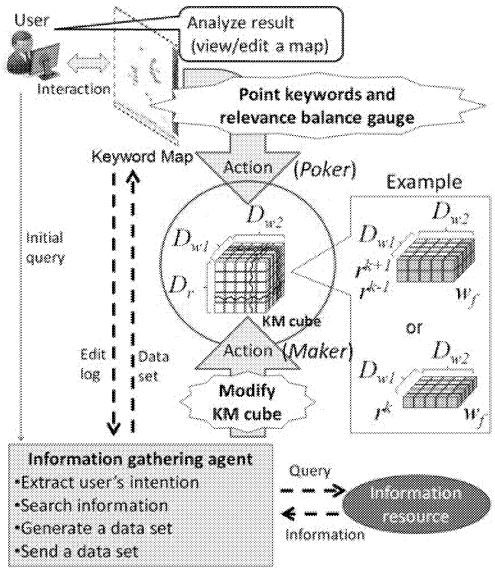


図 4 Poker-Maker モデル

ユーザは，初期クエリを情報収集エージェントに入力し，検索結果から構築される初期状態のKMキューブに基づくキーワードマップを得る．ユーザはインタラクション機能を用いて注目キーワードを設定したり，関連バランス制御を行い様々な視点からデータを眺めたりする．インタラクション機能を用いたデータ分析は，4.2節で述べたKMキューブへの操作に対応する．ユーザが，さらなる情報検索を望みキーワードマップで編集終了ボタンを押すと，ユーザによって指示された

注目/削除キーワード集合と関連バランス設定が，情報収集エージェントへ送られる．

情報収集エージェントは，ユーザの検索意図に応じて，仮想的な多次元データキューブで表される情報空間から，キーワード集合S_wを抽出し，D_{w1}，D_{w2}を構成する要素を決定する．関連属性についても，D_rを構成する要素を決定する．これらの要素についてダイシングを行い，KMキューブを再構築し，4.4節で述べた手法によりキーワードマップ用データセットを作成することで，ユーザの検索意図に合わせて修正された検索結果の可視化が可能になる．基本的に，再構築されたKMキューブには，注目キーワードは残るが，他のキーワードは検索結果に応じて変化する．図4では，注目キーワードW_fを含んだD_{w1}，D_{w2}によりKMキューブを構築している（図中のExample）．情報収集エージェントのアルゴリズムやタスク次第で，再構築前のKMキューブに存在したキーワードを全く残さないことや，反対に，構築前後で複数の重複するキーワードを複数存在させることもできる．

表2は，初期状態のKMキューブに基づく可視化が行われた後で，ユーザが実際に行う作業と，情報収集エージェントが実際に行う作業についてまとめたものである．初期クエリ入力後にユーザが操作するのは，あくまでキーワードマップで，ユーザにとって情報収集エージェントはブラックボックスとして扱われる．また，ユーザは，マップを編集して情報の分析作業を行うだけで，別途，検索クエリを生成するといった作業を意識的に行う必要はない．

表2 検索におけるユーザとエージェントの操作

User	・View result on the Keyword Map ・Edit a map (Point keywords and relevance balance setting)
Agent	・Extract user's intention from a edit log ・Search information according to user's intention ・Modify KM cube ・Generate a data set for the Keyword Map ・Send a data set

6 提案モデルによる既存システムの理解

提案モデルの意義は，意思決定のための探索的情報検索支援システムの設計に役立つことと，既存システムの理解に役立つことである．本節では，提案モデルに基づく探索的情報検索支援システムについて，3種類の既存システムを取り上げ，提案モデルによりその機能を整理可能であることを示す．また，不足する機能についても議論することにより，提案モデルが既存システムの理解に役立つことを示す．3つの既存システムそれぞれで想定されているユーザのタスク型（Task type）は異なり，本稿では，「話題検索型（Topic search type）」，「伝搬分析型（Distribution analysis type）」，「代替案選択型（Alternative selection type）」と呼ぶ．話題検索型は，ユーザの興味のある話題に関連する情報を収集するタスクである．伝搬分析型は，Web上の話題の伝搬から，話題の発信元となった影響力の高い記事の発見を行うタスクである．代替案選択型は，多数の代替案の中からユーザの希望に合わせたより良い選択を行うタスクである．提案モデルの妥当性を異なるタスクで実証することで，提案モデルがタスクに依存しない高い汎用性を備えていることを示す．以下では，4.2節で述べた意図との対応の観点から，既存システムの概要を説明する．

6.1 話題検索型タスクの支援

このシステムは，ユーザが興味をもったキーワードに関する情報を広く収集する作業を支援するものである[20]．この研究で定義しているキーワードと関連属性は以下のとおりである．

キーワード:収集した文書から抽出した単語．
関連属性：単語の共起度．

ユーザは，可視化されたキーワードの中から興味をもったものを選択し，それに関連する単語の検索を行うプロセスを繰り返す．これにより，単語をクエリとする一般的な検索エンジンを用いた場合に課題となる，ユーザが適切なクエリを思いつけず情報の絞り込みが困難になる状況や，逆に，話題に関連する単語がわからず，収集する情報が不十分になる状況を解決できる．このタスクでは，ユーザはキーワードマップ上で興味あるキーワードを注目キーワードとして（複数可）選択し，情報エージェントはそれをクエリとして関連キーワードを検索する．これは4.2節の検索意図K1に対応する．また，本タスクで用いる関連属性は1種類であるため，関連属性に関する検索意図は考慮されていない．

本タスクにおけるデモンストレーションとして，具体的には，キーワード「人工知能」を初期クエリとしたWeb検索結果から初期KMキューブを作成している．ユーザは，可視化されたキーワード「ロボット」「論理」（これら2つのキーワードの間には直接関連は無い）を注目キーワードとし（ダイシングを行い）再検索した．その結果，「ロボット」と「論理」が直接関連するKMキューブが構築されている．

6.2 伝搬分析型タスク支援

このシステムは，Blogを対象にし，ニュース伝搬の様子から話題の中心を探索する作業を支援するものである[21][22]．この研究で定義しているキーワードと関連属性は以下のとおりである．

キーワード: ニュース記事，エントリ，Blogサイト．
関連属性: 上記3 種類の組み合わせに関する計6種類．

Blog空間の構成要素（キーワード）をニュース記事，エントリ，Blogサイトと定義し，エントリにおけるニュース記事の引用や，エントリとそれを投稿したBlogサイト（ブロガー）との関係などに基づき，3 種類のキーワードの組み合せ，計6 種類についてそれぞれ関連属性（基礎関連属性）を定義する．キーワードマップにおいて，これら6 種の基礎関連属性全てを考慮して分析することはユーザにとって負担となるため，ニュース伝搬分析の観点から必要となる以下の3種類の中心関連属性に集約し，キーワードマップ用のデータとして用いる．

ニュース記事中心関連属性： ニュース記事間についての基礎関連属性．
エントリ中心関連属性： エントリ間，エントリ- ニュース記事，エントリ- Blog サイトの3 種類の基礎関連属性を合成．
Blog サイト中心関連属性: Blog サイト間，Blog サイト- ニュース記事の2 種類の基礎関連属性を合成．

本タスクにおいて，情報収集エージェントは，意図強調機能が適用されたキーワードをクエリとして検索を行い，中心関連属性について関連するキーワードを収集する．収集されたキーワードに，意図強調機能を適用されていない注目キーワード，および現在のマップにおいてそれらと関連をもつキーワードを加え，削除キーワードを除去する事により，

キーワードマップが再構築される．これは，検索意図K1，K2に対応する．また，検索時には6種類の基礎関連属性を用いて関連オブジェクトを収集し，情報収集エージェントで上記3種類の中心関連属性に集約する．これは，KM キューブ再構築時にロールアップを行っていることに対応する．

本タスクにおいて，ユーザは，話題となっているBlogサイトを探す際，一般的なWebブラウザで確認が困難かつ重要な手掛かりであるBlogサイト間の関連に着目し，Blog サイト中心関連属性でKMキューブをスライスした結果から，Blogサイトが注目されているかどうか判断している．

6.3 代替案選択型タスク支援

このシステムは，オンラインショッピングを対象にし，相補/非相補的な意思決定方略[19]に対応した，商品の比較検討作業を支援するものであり[17][23]，購買のために重要な外的情報検索[24]意欲の向上を目的としている．この研究で定義しているキーワードと関連属性は以下のとおりである．

キーワード：映画DVD タイトル．

関連属性：主演，ジャンル，監督．

情報収集エージェントはAmazonAPIを用いて実装されており，アイテムに関する情報はAmazonのサイトから収集する．キーワードマップにはDVDタイトルしか表示されないが，タイトルの上にマウスカーソルを置くことにより，商品の詳細情報（具体的な主演者名など）を別ウィンドウで確認できる．情報収集エージェントは，注目キーワードと共通する属性をもつDVDを検索し，KMキューブを再構築するが，極小化された関連属性は検索に用いない．また，関連属性が傾斜化された場合は，より重みの大きい属性が注目キーワードと共通しているDVDを重点的に検索する．本タスクの特徴は，関連する商品アイテムを幅広く確認する発散的過程だけでなく，購買する商品を絞り込んでいく収束的過程も含むことである．そのため，検索意図K1だけでなく，絞り込みに関連するR1，R2，R3 も考慮している．

本タスクにおいて，ユーザは，自身が採用した意思決定方略に基づき頻繁にデータ分析・再検索を行っているが，編集終了ボタンを押してから新しいキーワード配置を得るまでの時間は秒単位（長くて10秒程度）であるため，対話的なデータ分析操作と合わさり比較的スムーズな検索が行える．また，ユーザは既知アイテムに頼った検索だけでなく，提示された未知のアイテムに着目し，未知情報を幅広く探索できている[17][23]．

本タスクにおいて，ユーザは，キーワードマップの編集（KMキューブをスライシング，ダイシング，演算）による情報分析作業を通して，今まであまり興味の無かった映画へ興味を持つ場合がある．そして，興味を持った映画と同じ特定俳優や特定ジャンルの映画に関してさらなる情報収集を行うべく，演算映画タイトルの選択や関連バランス制御を行って再検索する．再検索後のKMキューブを構成するキーワード集合には，特定俳優や特定ジャンルに一致する映画タイトルが，再検索前と比較して数多く含まれる．

6.4 議論

本節では，前節で述べた3 つの既存システムを，提案モデルに基づきユーザの検索意図とそれに応じたKM キューブに対するアクションについて分析することで，提案モデルにより既存システムの機能を整理可能であることを示す．また，既存システムに不足する

機能についても議論する.

6.4.1 検索意図とアクション

6.1〜6.3節で概説した,タスクの異なる研究事例において,タスク型(異なる分析意図をもつ),ユーザのKMキューブに対するアクション,情報収集エージェントのKMキューブに対するアクションをまとめると,表3のようになる.

表3 ユーザによるKMキューブへのアクションとエージェントによるKMキューブへのアクション

Task types	User (search intention)	Agent
Topic search	Slice (K1)	Dice
Distribution analysis	Slice (K1), Dice (K2)	Dice, Roll-up
Alternative selection	Slice (K1, R1), Dice (R2), Calculation (R3)	Dice

表3は,ユーザと情報収集エージェントが共に,KMキューブに対するOLAPの基本操作で,ユーザの検索意図を取り扱えていることを示している.したがって,タスクが異なっても,ユーザの検索意図を,KMキューブに対するアクションとして統一的に扱えることがわかる.情報収集エージェントは,タスク毎に異なるユーザの多様な検索意図に対し,ダイシング・ロールアップにより,仮想的な多次元キューブである情報空間からKM キューブを再構築できている.したがって,ユーザ・情報収集エージェントともに,OLAP 操作に基づいて統一的にKMキューブを取り扱うことで,検索意図の異なる様々なタスクに柔軟に対応できることがわかる.また,再検索を行う度に,ユーザの検索意図に応じて可視化空間に対応するKM キューブが動的に変化していくことで,ユーザの探索空間が移動していき,探索的情報検索が支援可能であると考える.

既存システムでは,情報収集エージェントによる関連属性のドリルダウンは導入されていないが,ユーザの検索意図R1(ある関連属性の極大化操作)に対して,関連属性を細分化してKM キューブを再構築することも可能と考える.可視化の際,細分化した関連属性の重みを全て等しくすれば,細分化せず1つの関連属性からデータを眺めるのと同じことになるため,ユーザの検索意図を尊重しつつ,細分化した関連属性を提示することで,さらに異なった視点からのデータ分析手段を提供できると考える.この場合でも,キーワードマップから情報収集エージェントへ伝達する情報は,ユーザの指示したキーワード集合と関連バランス設定のみである.よって,提案モデルは将来への拡張性も備えていると考える.

6.4.2 既存システムで不足している機能

提案モデルと照合すると,6.1節の事例で用いられたシステムには,関連属性が1種類しかなく,複数の関連属性から特定の関連属性に着目する意図R1〜R3(4.2節)が反映できない.ユーザによって特定の関連属性を指示する機能が不足しているため,ユーザは,単語の共起度以外のどのような属性でキーワード同士が関連しているのか分析できず,また自身が重要視する属性を情報収集エージェントへ伝達できない.

提案モデルに基づき,複数の関連属性を扱い特定の関連属性を指示する機能を加えることで,情報収集エージェントが特定の関連属性に着目するユーザ意図を抽出し,それに応じた検索を行い,その結果,よりユーザの要求に適った検索結果を提示できると考える.

7 おわりに

本稿では，探索的情報検索を伴う多様なデータ分析タスクを支援するために，対話的な汎用的情報可視化ツールであるキーワードマップとエージェントアプローチを統合したシステムアーキテクチャを提案した．

ユーザの分析意図は，KMキューブに対するスライシング・ダイシングで表現でき，検索意図は，キーワードマップ上で着目したキーワード集合と関連バランス設定を指示する操作で表現できる．

ユーザによるデータ分析操作から対応する検索意図を抽出して情報収集エージェントへ伝達し，情報収集エージェントが，ユーザの検索意図に応じてKMキューブを再構成していく枠組みを，Poker- Maker（POint KEywords and Relevance balance setting - Modify Active Keyword map cube Expressing Relevance data）モデルとして提案した．提案モデルに基づいた，既存システムの理解から，探索的情報検索を伴う多様なデータ分析タスクにおいて，提案モデルが有効に機能したことを示した．

今後の研究の方向性としては，タスクごとに最適な（関連属性のロールアップ，ドリルダウンも考慮した）KMキューブを構成するアルゴリズムの検討があげられる．

参考文献

[1] 河野浩之；山田誠二；北村泰彦；高橋克己：「第2章情報収集エージェント」，情報検索とエージェント，東京電機大学出版局，pp. 27-52，2002.

[2] K. Sycara: “In-context information management through adaptive collaboration of intelligent agents,” in Intelligent Information Agents: Cooperative, Rational and Adaptive Information Gathering on the Internet, Springer, pp.78-99, 1999.

[3] S. T. Card, J. D. Mackinlay, and B. Shneiderman (eds.): “Chapter 1 Information visualization,” in Reading in Information Visualization: Using Vision to Think, Morgan Kaufman, pp.1-34, 1999.

[4] 網谷重紀；堀浩一：「知識創造過程を支援するための方法とシステムの研究」，情処学論，Vol. 46，No. 1，pp. 89-102，2005.

[5] 宮寺庸造；田地晶；及部佳代子；横山節雄；近谷英昭；夜久竹夫：「学術論文関係情報のグラフ描画問題に基づく視覚化手法」，信学論(D)，Vol. J87-D-I, No. 3, pp. 398-415, 2004.

[6] 角康之；堀浩一；大須賀節雄：「テキストオブジェクトを空間配置することによる思考支援システム」，人工知能誌，Vol. 9, No. 1，pp. 139-147，1994.

[7] 梶並知記；山口亨；高間康史：「キーワードマップを用いたWeb情報検索インタフェースのためのキーワード配置支援手法の検討」，計測自動制御学会論文誌，Vol. 42, No. 4, pp. 319-326，2006.

[8] 梶並知記；高間康史：「ユーザ意図を強調したキーワード配置支援機能を備えたインタラクティブなキーワードマップ」，情処学論，Vol. 48，No. 3，pp. 1176-1185，2007.

[9] G. Chang, M. J. Healey, J. A. M. McHugh, and J. T. L.Wang: “Chapter 5 Data mining,” in Mining the World Wide Web ～An Information Search Approach～, Kluwer Academic,

pp.67-69, 2001.
［10］Nikitas-S. Koutsoukis, G. Mitra, and C. Lucas: “Adapting on-line analytical processing for decision modeling: the interaction of information and decision technologies,” Decision Support Systems, Vol.26, No.1, pp.1-30, 1999.
［11］H. Reiterer, G. Tullius, and T. M. Mann: “INSYDER: A content-based visual-information-seeking system for the Web,” International Journal on Digital Libraries, Vol.5, No.1, pp.25-41, 2005.
［12］石谷康人；鈴木優；布目光生：「意味クラス解析と意図推定に基づくインタラクティブな情報検索インタフェース」，情処学論, Vol. 48, No. 12, pp. 3793-3808, 2007.
［13］山田誠二；大澤幸生：「WWW における概念理解のためのナビゲーションプラニング」，人工知能誌，Vol. 14，No. 6，pp. 1125-1133，1999.
［14］山田信太郎；松永吉広；伊東栄典；廣川佐千男：「Web シラバス情報収集エージェントの試作」, 信学論 (D), Vol. J86-D-I, No. 8，pp. 566-574，2003.
［15］V. Lesser, B. Horling, F. Klassner, A. Raja, T. Wagner, and S. XQ. Zhang: “BIG: An agent for resource-bounded information gathering and decision making,” Artifical Intelligence, Vol.118, pp.197-244, 2000.
［16］阪本俊樹；北村泰彦；辰巳昭治：「競争型情報推薦システムとその合理的推薦手法」，信学論（D），Vol. J-86-D-I，No. 8，pp. 608-617，2003.
［17］梶並知記；槇原崇；小笠原敏之；高間康史：「関連バランス制御機能を組み込んだキーワードマップによる意思決定方略に応じたデータ分析の支援」，知能と情報，Vol. 21，No. 6，pp. 1067-1077，2009.
［18］C. Chen: “Chapter 3 Graph drawing algorithm,” in Information Visualization Beyond the Horizon (2nd.ed), Springer,pp.65-88, 2006.
［19］R. M. Hogarth: “Chapter 4 Combining information for evaluation and choice,” in Judgement and Choice, John Wiley, pp.62-85, 1987.
［20］Y. Takama and T. Kajinami: “Relevance feedback based on interactive keyword map system,” 8th International Conference on Control, Automation, Robotics and Vision, pp.1814-1819, 2004.
［21］Y. Takama, A. Matsumura, and T. Kajinami: “Visualization of news distribution in Blog space,” Intelligent Web Interaction 2006 (WI-IAT’06 Workshop), pp.413-416, 2006.
［22］Y. Takama, A. Matsumura, and T. Kajinami: “Interactive visualization of news distribution in Blog space,” New Generation Computing, Vol.26, No.1, pp.23-38, 2008.
［23］T. Kajinami, J. Komiya, T. Ogasawara and Y. Takama: “Application of keyword map to decision support through exploratory search,” 2008 IEEE International Conference on Systems, Man and Cybernetics, pp.2177-2181, 2008.
［24］杉本徹雄（編）：「4 章 消費者の情報探索と選択肢評価」，消費者理解のための心理学，福村出版，pp. 56-72，1997.

(2010 年 4 月 9 日受付)
(2010 年 8 月 21 日採択)
(2010 年 9 月 10 日オンライン公開)

論文

漸近的対応語彙推定法に基づく翻訳文の解釈的特徴の抽出 −日本語翻訳聖書の計量的比較−

Extracting the interpretive characteristics of translations based on the asymptotic correspondence vocabulary presumption method: Quantitative comparisons of Japanese translations of the Bible

村井　源 [1*]
Hajime Murai

*1　東京工業大学社会理工学研究科価値システム専攻
Tokyo Institute of Technology, Graduate School of Decision Science and Technology
h_murai@valdes.titech.ac.jp

翻訳の背景にある解釈については古くから翻訳の比較や背景的な思想の分析が行われてきたが，人文学的手法によるものであるため客観性の担保が困難で，大規模な比較分析には適さなかった．本研究では情報技術を活用し，計量的な翻訳の比較と背景的思想の特徴抽出を行う手法を提案する．まず，原文と翻訳文の対応語彙の特定・抽出アルゴリズムを考案した．単語の一対一対応の仮定，被特定単語の除去による相互情報量の再計算，閾値の漸近的低減を組み合わせることで従来の手法に比べて再現率が 20%程度向上し，一つの単語の複数の単語への対応も抽出可能となった．次に，語彙の対応関係を用いて二種類のネットワークを作成し中心性の計算から各翻訳での特徴的な概念を示す単語を抽出した．抽出結果は，翻訳作成の背景的な思想と合致しており，分析手法の有効性が確認された．

Although there have been some studies of translations that focus on background interpretations by comparing and analyzing translations, such studies have utilized the methodologies of the humanities, which are not suitable either for maintaining objectivity or for large-scale analysis. Utilizing information technologies, this paper proposes some methods for numerical comparisons and the extraction of background interpretations for translations. Firstly, a new algorithm is designed to identify word pairs between the original text and the translated version. The algorithm incorporates three features; namely, a word-for-word correspondence hypothesis, a recalculation of mutual information after the elimination of identified pairs and an asymptotic threshold reduction. Through the combination of these features, recall rates improve by 20% over conventional methods, and it is possible to extract multiple words corresponding to each word in the word pairs. At the next stage, two types of networks are created based on the word correspondences, and the characteristics of translated words are extracted by calculating centrality values. The extracted results correspond to the background thoughts in translation making, and so confirm the validity of this method.

キーワード：翻訳，解釈，語彙の意味推定，translation, interpretation, word sense disambiguation

1. はじめに

歴史を通じ多大な影響を与えてきた古典的テキストにおいては，時代や社会の変化に応じて様々な形での現代語への翻訳が幾度となく試みられてきている．また同時代にあっても，古典的テキストに対する解釈の相違に基づいた，複数の異なる翻訳がされる場合は少なくない．これら翻訳の相違には，古典的テキストへの理解の相違があると考えられており，翻訳者たちのテキスト理解と思想的特徴の抽出を目的とした研究が，人文学的な手法に基づいて様々に行われている[1]．しかし，人文学的な手法における分析は人手によるため，分析者の背景知識に基

づく深い洞察が可能である半面，大量のデータに対して一様な基準での分析を行うことは困難であり，結果の客観性も科学的な手法に比べ高くない．

一方で，情報学の分野では近年急速に発達してきた情報処理技術を用いた自動翻訳の研究が言語の統計的モデルに基づいて精力的に進められてきている[2]．自動翻訳においては，語彙の対応辞書や，大規模な翻訳例文のコーパスなどをデータとして用いる手法が一般的であるため，自動翻訳のための精度の良い大規模な語彙の対応関係の抽出や，対応関係にある翻訳文コーパスを作成する技術が盛んに研究されてきた[3][4]．さらに，これらの結果を向上させるために，非対応関係のコーパス[5]や，専門分野のテキストに絞り語彙の対応関係を抽出する手法[6]などさまざまな発展的研究もなされてきている．

また，これらの出力として得られる翻訳文の自然言語としての適切さをどのように実現するかは大きな課題となっており，人間が翻訳した場合と比較して翻訳文の自然さを評価する計量的な指標なども提案されている[7]が，これらの手法では人間が翻訳したテキストが適切には評価されない問題点も指摘されている[8]．

本論文では，これらの研究を踏まえ，情報処理技術に基づいた，計量的かつより客観的な翻訳の比較手法の可能性を探る．計量的な翻訳比較には様々な手法の可能性が考えられるが，本論文では語彙のレベルに対象を絞る．具体的な手法としては，語彙の原文と翻訳文の対応関係を，計量的データに基づいて推定し，各翻訳での語彙の対応関係の相違から，各翻訳者の背後にある原文への解釈を客観的・計量的な形で抽出することを目的とする．

2. 対象となるデータ

2.1 テキスト

本論文では，計量的翻訳比較のための対象データとして，日本語訳新約聖書のテキストを用いる．新約聖書は古代より世界中で数多くの言語に翻訳されてきた歴史を持ち，現代にいたってもなお多くの翻訳プロジェクトが進行中である．

日本語でも戦国時代より様々な翻訳が試みられてきたが，現在広く用いられている翻訳は，口語訳[9]，新改訳[10]，新共同訳[11]の三種類である[12]．口語訳聖書はプロテスタントの聖書学者たちによって 1955 年に完成した初の本格的口語体聖書で，現在でもプロテスタント教会の礼拝等で用いられる．口語訳は，歴史学・考古学の知見や，写本比較による本文批判などを用いる近代的な聖書批評学の影響を受けて作られた．また，当時の日本語改革の影響を受け，読みやすさが重視されている．これに対して神やキリストの権威が弱められていると反発した福音派の人々が 1970 年に新改訳を出版した[13]．福音主義では聖書を誤りなき神の言葉として解釈を行うため新改訳では原文に忠実に逐語的に翻訳することが方針とされた．

これらの聖書はプロテスタント学者達によるものであったが，キリスト教各教派の合同を目指したエキュメニズム運動が 1960 年代以降カトリックでも盛んになり，日本でもカトリックとプロテスタントの合同の翻訳が共同訳として 1978 年に出された．しかしカトリックとプロテスタント双方に遠慮して妥協的翻訳を行ったためさまざまな批判を浴び，問題点の解消を図って 1987 年に新共同訳として新たにされた．現在この新共同訳聖書が最も一般的に用いられている聖書翻訳である．

本論文では，JBible[14]所収の三翻訳のテキストデータを分析に用いた．また，新約聖書のギリシア語原文としては世界的に新約聖書の底本として用いられる，Nestle-Aland[15]を選択した．ギリシア語テキストデータは BibleWorks[16]所収のものを用いた．

2.2 文のアラインメント

複数の言語間での語彙の対応関係の分析では，原文と翻訳文間で文単位による単語の出現頻度を用いるのが一般的である．しかし，作者による文の切れ目が明確である現代のテキストと異なり，古代のテキストにおいては文の切れ目が明確ではない．このため，翻訳ごとに解釈が異なり，文の区切り目に相違が出る．このため，本論文では，聖書学で一般的に用いられる章と節をテキスト区

切りとして用いる. 節は2～3 文程度を一つにまとめた単位であり, 新約聖書においては, 全 27 テキストに章が 260 ありその中に 7928 節が含まれる. なお, 節の数は写本の解釈で異なりうるが, 本論文は原文の底本である Nestle-Aland[15]に従う.

2.3 語彙

本論文のテキストの分析単位は語彙であるが, 新約聖書ギリシア語テキストの形態素分割データとして BibleWorks[16]に付属の語彙分析結果を用いた. また日本語の各翻訳文の形態素解析には TextSeer[17]を用い, 聖書関連用語辞書を通常の日本語辞書に追加して形態素解析を行った.

本論文では, 概念関係の翻訳に焦点を絞り, ギリシア語の前置詞, 接続詞等と日本語の助詞, 助動詞等の非内容的品詞を削除し, 名詞・形容詞・動詞・副詞・連体詞の自立語を対象とした. これらの品詞中でも, 受け身を表す「れる・られる」, 敬意を表す「御・み・お・様」, 複数を表わす「達・たち・ども」はギリシア語では対応単語がないか, あるいは語の屈折となり一つの単語とはならないため, 本論文では分析対象から除外している.

表 1 各テキストの分析対象単語数

		対象品詞の単語数	使用単語数	比率
Nestle Aland	単語	5,064	2,417	47.7%
	頻度	81,489	76,672	94.1%
口語訳	単語	6,305	4,198	66.6%
	頻度	99,524	97,417	97.9%
新改訳	単語	5,918	4,013	67.8%
	頻度	101,170	99,265	98.1%
新共同訳	単語	6,008	4,150	69.1%
	頻度	99,113	97,255	98.1%

また, 対応関係にある単語を抽出するためには, 単語の出現位置の類似性が指標となる. 出現頻度の少ない単語を分析に含めると, 分析結果の精度が著しく低下するし, 出現頻度の非常に高い単語のみを用いると推定される単語の数が低下する. 複数の閾値で試行した結果, 精度と推定語彙数のバランスを取って, 原文での出現頻度3未満の単語及び, 翻訳文の出現頻度2未満の単語は対象から除外した. 分析対象の品詞に属する全単語数と, その中から頻度の条件に適合した分析対象単語数を表 1 に示す. 表 1 より, 分析対象単語種類数では全体の 4～6 割程度と若干低いが, 単語の出現頻度では 94%～98%をカバーしている.

3. 漸近的対応語彙推定法

3.1 語彙推定と単語出現確率

翻訳関係にあるテキスト間で, 語彙の対応関係を自動抽出する場合, 再現率(Recall)と精度(Precision)が問題となる. 語彙の対応関係における再現率とは, テキスト中の全単語中(図 1 中の Tall)で正しい対応関係が抽出された単語(図 1 中の Tright)の比率である. また, 精度とは抽出された全対応関係(図 1 中の Tguess)中で正しい対応関係の単語の比率である. すなわち、再現率=Tright／Tall であり, 精度= Tright／Tguess である. ただし, 単位を単語の種類の数とするか, テキスト中での単語の出現頻度の合計とするかで結果の値は異なる. なお, 本論文での再現率の定義は翻訳テキスト間での語彙の対応分析において用いられている用法に基づいており, 情報検索における定義とは異なる.

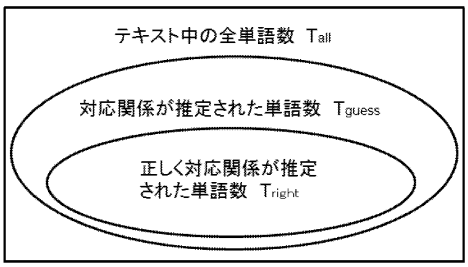


図 1 再現率と精度の関係

従来の情報学的諸研究においては, 翻訳と原文での語彙の対応関係を抽出する目的は, 自動翻訳での利用が主であった. このため, 原文と翻訳間での再現率は重視されず, 精度が高いことが求められてきた[6,18]. これは精度が十分な値ならば, テキストの規模を拡大するだけで, 容易に自動翻訳辞書として十分な数量の対応単語の組を抽出することが可能となるからである.

しかし, 本研究のような翻訳文間の比較を行う場合には, テキスト中の出現単語の再現

率が低い対応語彙推定法では信頼性の高い計量的比較を行うことができない．そのため情報学で用いられてきた従来の対応語彙推定のアルゴリズムを修正し，推定結果の再現率を上昇させる必要がある．

さらに，従来の手法においては，自動翻訳で用いることのできる最適な語彙の対応関係を抽出することが主眼であったため，ある単語が複数の単語に訳し分けられている場合も最も翻訳される確率の高い単語を特定することが目的とされてきた[2]．しかし，各翻訳における解釈の相違を分析する場合には，一つの単語が，異なる複数の単語へどのように訳し分けられているかという点も重要であると考えられる．

一方で，語彙の自動意味分析の分野では，複数の語義を持つ単語の使用例を大規模な対訳コーパスから計量的に抽出し，自動的な意味タグ付けに用いることも行われている[19]．しかしこれらの研究においても，各語義の使われ方の特徴抽出の個数と精度が重要であり，原文と翻訳での語彙の対応関係の再現率の向上は目的とされていない．

また，古典語で記されたテキストのような，同じ言語のテキストが容易に大量に入手可能でない対象の分析においては，他の大量のテキストを学習データや弁別用のデータとして用いる手法は適用困難である．さらに，本論文で扱うような古典テキストでは聖書のように文の切れ目が明確でない場合が少なくなく，この点においても従来の語彙対応関係推定手法よりもより精度の高い手法が求められる．

以上より，本論文で求められる語彙の対応関係推定アルゴリズムの備えるべき特徴は以下のようにまとめられる．

- 分析対象テキストで再現率が高い
- 複数単語への訳し分けに対応
- テキストの切れ目が文単位でなくても分析が可能
- 他の大規模なコーパスや大量の学習データを必要としない

3.2 対応語彙推定アルゴリズム

多くの語彙対応関係推定アルゴリズムにおいては，原文の単語 o と翻訳文の単語 t の間の関係性の指標として相互情報量が用いられる．相互情報量とは，原文で o が出現した場合に対応する翻訳文で t が出現する確率 P(t|o)と，逆に t が出現した場合に対応文で o が出現する確率 P(o|t)の積である．すなわち o と t の相互情報量 Iot は，$Iot = P(t|o) \times P(o|t)$ と定義できる．基本的には相互情報量が大きいほど，2 単語が翻訳での対応関係にある可能性が高いと推定される．

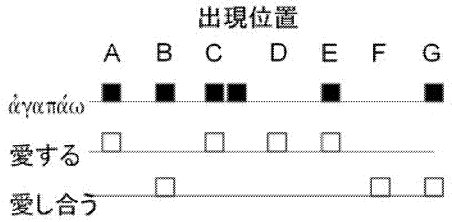


図 2 単語の対応する組の関係

図 2 に例として，ギリシア語の”ἀγαπάω”と日本語の「愛する」「愛し合う」での対応関係を示す．この場合”ἀγαπάω”は C で二回出現し，CDF 以外の箇所では一回ずつ出現している．”ἀγαπάω”と「愛する」は箇所 ACE で対応しており，これらより相互情報量は，3/6×3/4=0.375 となる．同様に”ἀγαπάω”と「愛し合う」の相互情報量は約 0.22 と計算できる．単語の組の抽出での相互情報量の閾値が 0.3 の場合，”ἀγαπάω”と「愛する」が対応する組として抽出され，ACE の3箇所で”ἀγαπάω”が合計3回推定されたことになる．閾値が 0.2 の場合には，ABCEG の 5 箇所で合計 5 回”ἀγαπάω”の対応単語組が推定され，推定単語の頻度合計は 5 である．

一般的にはこのように相互情報量で閾値を設け，閾値以上の相互情報量を持つ対応単語の組を抽出する．閾値が高いと対応関係が正しい確率，つまり精度は向上するが，閾値を満たす単語のペアが少なくなり再現率は低下する．基本的に再現率を上げたい場合には相互情報量の小さい単語対までを対応単語として抽出し，精度を上げたい場合には相互情報量の閾値を上げることになる．

また，閾値によって抽出される単語の組数と相互情報量の関係を図 3 に示す．図 3 では後述する漸近的対応語彙推定法による口

語訳での場合を示している. 図 3 より, 抽出組数は相互情報量に対し対数的な変化を示していることが分かる. このため, 以下での相互情報量の閾値は 0.1, 0.2, 0.4, 0.8…のような倍々の値を取ることとする。

相互情報量と精度・再現率の関係を例示するためにギリシア語聖書と口語訳聖書での相互情報量と精度・再現率の関係を表 2, 3 に示す.

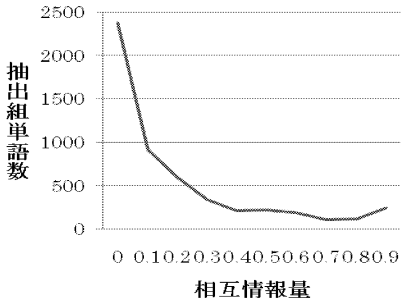


図 3 抽出組数と相互情報量の関係

表 2 最尤推定の精度(口語訳)

相互情報量	語数精度 Pn		頻度精度 Pf	
	全体	部分	全体	部分
0.64～	90.0%	90.0%	98.7%	98.7%
0.32～	90.9%	93.9%	98.0%	99.3%
0.16～	88.7%	90.6%	97.6%	98.7%
0.08～	84.4%	87.0%	96.4%	97.7%
0.04～	81.1%	83.3%	96.8%	97.6%
0.02～	79.6%	82.7%	96.5%	97.4%
0.01～	79.0%	83.0%	96.4%	97.4%

表 3 最尤推定の再現率(口語訳)

相互情報量	頻度再現率 Rf	
	全体	部分
0.64～	18.7%	18.7%
0.32～	36.7%	37.2%
0.16～	46.6%	47.1%
0.08～	49.6%	50.3%
0.04～	52.3%	52.8%
0.02～	52.8%	53.3%
0.01～	52.9%	53.5%

表 2,3 では, 一般的な最尤推定[2,18]に基づき, ギリシア語の各単語と最も高い相互情報量を持つ日本語をそれぞれ抽出した. その後相互情報量 0.01 以上の組からランダムに 100 組を選び, 人手で対応が正しいかを確認して精度・再現率の推定値を相互情報量の閾値ごとに算出した. ランダム選択された 100 組中, 相互情報量 0.64 を超える組のみの合否の合計が「0.64～」の行に記されている. 例えば「0.32～」は相互情報量 0.32 を超える組をすべて含むため, 「0.64～」に含まれる組も「0.32～」の中に含まれる. これは, 閾値を 0.32 にした場合に再現率・精度がどの値になるかを示すためである.

表 2 中の語数精度 Pn は正しい対応の単語の組数の精度を示し, 頻度精度 Pf は, 正しい対応が推定された全単語の出現箇所の合計による精度を示している.

なお語数精度 Pn の分母と分子の単位は図 1 での場合と異なり単語数ではなく単語組数となっている. これは従来の精度計算では最尤推定法を用いていたため一つの原語に対して推定される翻訳語は一つであったのに対して, 本論文では一つの原語に対して複数の翻訳語の対応を抽出しうる. このため, 単位が語数ではなく対応組数となっている.

原文中で分析対象となる全単語種類数を Tall, 対象単語の原文の全出現頻度の合計を Fall とし, 推定された原文単語中の単語組数を Tguess, 原文での推定単語の頻度合計を Fguess, 正しく推定された原文中の単語組数を Tright, 正しく推定された原文での推定単語の頻度合計を Fright とするとき,

$$Pn = Tright / Tguess$$
$$Pf = Fright / Fguess$$

と計算できる.

また再現率も同様に, 語数再現率 Rn と頻度再現率 Rf を

$$Rn = Tright / Tall$$
$$Rf = Fright / Fall$$

と計算できる. ただし, 語数再現率 Rn では, 分母が単語種類数(単位は語数)であるのに対し, 分子は抽出された単語組数となる. このため原文中の一つの単語が翻訳文中の複数の単語と対応組を作ることで, 理論上は語数再現率 Rn は 1.0 以上の値を取りうる. また, 対応の正しくない原語の対応組が多数あった場合も, 他の原語と多数種類の翻訳語が対応していれば再現率の値は高く算出される可能性があり, 抽出された単語組の

対応の正確さの指標としては不適切と考えられる. そこで以下の表では語数精度, 頻度精度, 頻度再現率のみを示す.

また, これらの計算においては原文と翻訳文中の単語の部分的な対応を含む場合と含まない場合が考えられる. 表 2 中での「全体」は単語の対応関係が一対一であり, 語義の全体が正しく対応する場合のみの再現率・精度を示す. 表 2 中の「部分」の項は原文の単語が翻訳文で複数の単語に分割して翻訳され, それらの単語の一つに適合した場合(ex. “σαλπίζω”が「ラッパを鳴らす」の「ラッパ」・「鳴らす」のどちらかに対応する)と, 逆に原文の複数の単語が翻訳文では一つの単語に翻訳されている単語に適合した場合(ex. “ϵ’τέραις”と“γλώσσαις”のどちらかが「異言」という一単語に対応)を含んで再現率・精度を計算した結果を示している.

表 2,3 より, 相互情報量が高い部分はデータ点数が少ないため値が多少上下するものの, 相互情報量の閾値を下げると精度が下がり再現率が上がる様子が見て取れる. また, 原文と翻訳文で単語数と概念の対応関係に相違のある単語(「全体」と「部分」項の差分)は数%程度であることが読み取れる.

最尤推定法では対応単語の組数の精度 80%程度を実現する場合, 再現率は 5 割強程度である. なお, 言語的に近縁関係で, 語彙も相互に重複している比率の高い英語とフランス語でも, 精度を重視すると再現率は 6 割程度である[18]. このため, 文法も語彙構成も大きく異なるギリシア語と日本語の場合それを下回る数値となるのは予測された結果である.

3.2.1 一対一を仮定した推定

本論文においては, 一つの単語がどのように複数の単語に訳し分けられているかも, 翻訳の背後にある解釈を計量的に分析する上では重要な情報である. このため, 相互情報量が最大の一つの単語の組のみ抽出する最尤推定法ではなく, 原文の単語に対して相互情報量が一定値以上の複数の翻訳語を抽出することを考える. ただしこの場合, 単純に相互情報量だけを用いると, 実際には対応関係にはない共起単語の訳語が多数抽出される懸念がある.

そこで表 2, 3 の原文の単語が分割や結合されて翻訳される場合に着目する. 相互情報量の閾値 0.01 の場合, 抽出された単語の組中でこれらの分割・結合を含む組は個数精度では 4.0%, 頻度精度で 1%, 頻度再現率で 0.6%と, 抽出された単語組の中では少数であることが分かる. この結果より語彙がテキスト全体で一対一対応であるとの仮定をおき, 相互情報量が高い単語から優先的に推定を行い, 後の推定ではすでに推定された出現位置と重なる単語は推定されないようにする.

単語の一対一対応の仮定に基づく, 単語の組の推定のプロセスを図 4 に例示する.

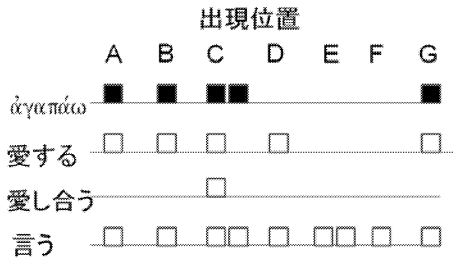


図 4 一対一対応仮定での推定1

まず, 図 4 で”ἀγαπάω”に対して候補となる翻訳語との相互情報量を計算する. このとき「愛する」とは 0.64, 「愛し合う」とは 0.2, 「言う」とは約 0.56 である. 単純に閾値で切ると, 0.6 以上なら”ἀγαπάω”と「愛する」の組のみが, 0.5 以上ならこれに合わせ”ἀγαπάω”と「言う」の組が対応組として抽出される.

これに対し, 一対一対応の仮定をおき, 閾値の高い物から順に推定していくと, 最初に”ἀγαπάω”と「愛する」の組が抽出される. 次に相互情報量が大きいのは「言う」であるが, これは「愛する」と出現位置が重複しており, 一対一対応仮定に反するので除外される. そして, 次に相互情報量の大きい「愛し合う」は出現位置が「愛する」と重複しない(C の位置は”ἀγαπάω”が 2 回出現しており, 2 回まで推定が可能である)ため対応組として推定される.

相互情報量の閾値が 0.2 以下ならば結果として”ἀγαπάω”の対応組は「愛する」と「愛し合う」と推定される. このように, 一対一の仮定をおくことで一つの原文の単語に対し複数の対応組が, 共起語を除外する形でより

正確に抽出されることが期待される.

上記のアルゴリズムで推定された相互情報量 0.01 以上の単語の対応に対して, 推定結果からランダムに 100 組を抽出し精度と再現率を計算した結果を表 4,5 に示す. 翻訳対象テキストには最尤推定の場合と同じ口語訳を用いた. また, 抽出結果と表 4,5 に基づき算出した抽出結果中の正答数の推定値を表 6 に示す.

表 6 中での累積語数精度・累積頻度精度は, ランダム抽出で計算された語数精度・頻度精度の比率と相互情報量の閾値以上の単語組数・推定単語の頻度合計との積である. 累積語数精度を An, 累積頻度精度を Af, ランダム計算による Pn と Pf の推定値をそれぞれ Pn', Pf'とすると,

$An = Pn' \times Tguess$

$Af = Pf' \times Fguess$

と定義できる.

表 4,5,6 の結果より, 相互情報量 0.01 では, 再現率が 10%以上向上しているが, 精度は若干下がっている. 頻度の低い用法も対象に含めたことで, 再現率は上昇したが, 精度は低下したと考えられる.

表 4 一対一仮定による精度(口語訳)

相互情報量	語数精度 Pn		頻度精度 Pf	
	全体	部分	全体	部分
0.64～	100.0%	100.0%	100.0%	100.0%
0.32～	100.0%	100.0%	100.0%	100.0%
0.16～	91.9%	94.6%	99.0%	99.5%
0.08～	91.1%	94.6%	98.7%	99.3%
0.04～	87.0%	92.8%	97.5%	98.9%
0.02～	79.3%	85.4%	95.1%	96.6%
0.01～	73.0%	79.0%	92.5%	94.6%

表 5 一対一仮定による再現率(口語訳)

相互情報量	頻度再現率 Rf	
	全体	部分
0.64～	18.6%	18.6%
0.32～	35.8%	35.8%
0.16～	47.4%	47.6%
0.08～	52.2%	52.5%
0.04～	57.7%	58.5%
0.02～	60.2%	61.1%
0.01～	63.2%	64.6%

表 6 一対一仮定による推定結果(口語訳)

相互情報量	累積語数精度 An		累積頻度精度 Af	
	全体	部分	全体	部分
0.64～	438	438	14,261	14,261
0.32～	858	858	27,453	27,453
0.16～	1,275	1,312	36,319	36,486
0.08～	1,749	1,817	40,111	40,296
0.04～	2,177	2,323	44,211	44,862
0.02～	2,512	2,705	46,139	46,879
0.01～	2,914	3,154	48,463	49,527

3.2.2 相互情報量再計算による推定

先のアルゴリズムで再現率が 10%程度向上したが, まだ全体の 2/3 であり十分とは言い難い. そこで, 先の一対一仮定での推定による被推定単語の使用箇所を除いた残渣データに対して再度相互情報量の計算を行う. このようにすると, 被推定単語部分を除外することで, 推定された単語の出現頻度の総数が低下し, 相対的に相互情報量が上昇する. 再計算によって, 相互情報量が閾値を超えた単語の組を再度抽出することにより再現率を上昇させられると期待される.

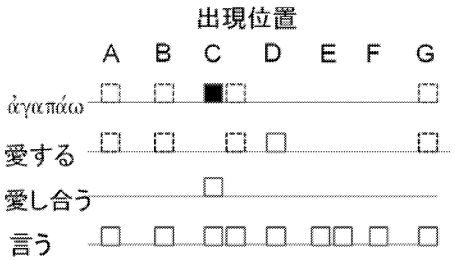


図 5 一対一対応仮定での推定2

これを先ほどの例で示すと, 図 4 での「愛する」と"ἀγαπάω"の対応単語の組が推定された後に, "ἀγαπάω"と「愛する」の対応ですでに使われた箇所を削除して相互情報量を再計算する. 図 5 での点線部分が最初の推定で対応組が特定された位置である. これらを除いて相互情報量を再計算すると,「愛する」とは 0,「愛し合う」とは 1.0,「言う」とは約 0.1 となり,「愛し合う」と"ἀγαπάω"の組の相互情報量が大幅に上昇していることが分かる. 一対一対応の仮定のみでも相互情報量の閾値を下げれば「愛し合う」と"ἀγαπάω"の組

が抽出可能であるが，閾値を下げることで妥当な組み合わせでない可能性の高い組も同時に抽出することになり，全体での精度が低下する懸念があった．しかし，相互情報量の再計算によって，相互情報量の閾値を高く保持したまま，複数語義の抽出が可能となる．

この仮定により，複数翻訳語との対応や，頻度の少ない単語との対応関係も高い相互情報量の組として抽出が容易となると期待される．

相互情報量の再計算を，閾値を超える単語の組が出現しなくなるまで繰り返した結果の精度・再現率を表 7,8 に，単語の組の抽出結果と表 7,8 に基づき算出した抽出結果中の正答の推定値を表 9 に示す．分析対象として用いた翻訳文は口語訳であり，精度・再現率の計算には先と同様にランダムに抽出した 100 組の単語の対応を用いた．

表 7,8,9 より抽出後のデータの再計算の結果再現率はさらに 10%程度上昇しているものの，精度はさらに減少しており，減少量は最尤法の場合に比べて各種類で数%から 10%程度であることが分かる．

表 7 再計算推定による精度（口語訳）

相互情報量	語数精度 Pn		頻度精度 Pf	
	全体	部分	全体	部分
0.64～	100.0%	100.0%	100.0%	100.0%
0.32～	91.7%	100.0%	97.8%	100.0%
0.16～	88.0%	96.0%	96.8%	98.7%
0.08～	87.5%	95.0%	93.3%	95.5%
0.04～	86.2%	93.1%	92.1%	94.5%
0.02～	82.7%	89.3%	91.5%	94.0%
0.01～	74.0%	82.0%	84.8%	91.1%

表 8 再計算推定による再現率（口語訳）

相互情報量	頻度再現率 Rf	
	全体	部分
0.64～	18.8%	18.8%
0.32～	37.5%	38.4%
0.16～	49.6%	50.6%
0.08～	56.4%	57.8%
0.04～	63.2%	64.8%
0.02～	69.7%	71.6%
0.01～	71.8%	77.0%

表 9 再計算推定による推定結果（口語訳）

相互情報量	累積語数精度 An		累積頻度精度 Af	
	全体	部分	全体	部分
0.64～	440	440	14,427	14,427
0.32～	819	893	28,746	29,407
0.16～	1,306	1,425	38,029	38,777
0.08～	1,847	2,005	44,910	45,793
0.04～	2,470	2,667	48,440	49,670
0.02～	3,109	3,360	53,453	54,903
0.01～	3,750	4,155	55,013	59,066

3.2.3 相互情報量漸近による推定

再現率は本論文の目的通り向上したが，非最尤の単語の対応や相互情報量の少ない単語の対応の抽出により低下した精度を向上させる必要がある．

精度低下の原因として考えられるのは，原文と翻訳文での単語の一対一対応の仮定をおき，被推定箇所から該当単語を削除して相互情報量の再計算を行うため，一度推定を誤った場合，それ以降の当該単語の推定成功の可能性は著しく減少することである．すなわち，一対一の仮定と繰り返し計算によって，次第に誤差が蓄積する構造に精度低下の原因があると推測される．そこで，プロセス上先に推定される単語の対応での誤推定率を減少させるため，相互情報量が高い，すなわち精度も高い単語の対応を優先的に推定することで，計算プロセス全体を通じて蓄積する誤差を低減することを考える．

具体的には，第一段階では相互情報量の閾値を高く設定し，100%に近い精度の単語の対応を一対一の仮定をおいて再計算により可能な限り抽出する．次に閾値を少し下げ，同様のプロセスで多少精度は下がるが比較的誤推定の確率の僅少な単語の組を抽出し尽くす．以降，この閾値の段階的な引き下げを繰り返し，早期の段階での精度を可能な限り上昇させることで誤差の蓄積を減らす．

以上のアルゴリズムを，相互情報量の初期閾値 1.0，ステップごとに閾値を 0.8 倍する形で相互情報量の下限を 0.005 として実行した場合の精度・再現率を表 10, 11 に示す．また，単語の対応組の抽出結果と表 10,11 に基づき算出した抽出結果の正答の推定値を表 12 に示す．他の表と同様に精度・再現

率の計算にランダムに抽出した 100 組の単語の対応組を用いた. 対象の翻訳文は口語訳である.

表 10 漸近推定による精度(口語訳)

相互情報量	語数精度 Pn		頻度精度 Pf	
	全体	部分	全体	部分
0.64～	84.6%	92.3%	98.4%	99.5%
0.32～	92.0%	96.0%	99.3%	99.8%
0.16～	85.0%	95.0%	98.1%	99.6%
0.08～	87.1%	95.2%	97.9%	99.5%
0.04～	85.5%	94.2%	97.5%	99.4%
0.02～	84.8%	93.7%	97.2%	99.2%
0.01～	80.4%	89.1%	95.6%	97.7%
0.005～	76.0%	85.0%	94.3%	96.8%

表 11 漸近推定による再現率(口語訳)

相互情報量	頻度再現率 Rf	
	全体	部分
0.64～	18.8%	19.0%
0.32～	41.0%	41.2%
0.16～	53.0%	53.9%
0.08～	65.5%	66.6%
0.04～	71.3%	72.6%
0.02～	76.5%	78.1%
0.01～	78.0%	79.6%
0.005～	79.5%	81.6%

表 12 漸近推定による推定結果(口語訳)

相互情報量	累積語数精度 An		累積頻度精度 Af	
	全体	部分	全体	部分
0.64～	446	486	14,398	14,570
0.32～	1,146	1,196	31,406	31,570
0.16～	1,804	2,016	40,670	41,305
0.08～	2,690	2,940	50,251	51,058
0.04～	3,280	3,614	54,648	55,696
0.02～	3,743	4,134	58,645	59,843
0.01～	3,884	4,304	59,766	61,066
0.005～	4,018	4,494	60,983	62,593

表10, 11より, 一対一の仮定をおいて被推定箇所の単語を除き, 相互情報量の再計算を繰り返し, 閾値を漸減させることによって, 精度を最尤推定と同程度に保ったまま再現率を 30%弱も向上させることに成功したことが分かる. この手法は他の大規模なテキストや学習データを必要とせず, 原文の一つの単語の翻訳での複数単語への訳し分けに対応しており, 本論文で必要な要件を満たしている.

この対応語彙の推定アルゴリズムを, 漸近的対応語彙推定法と名づける.

3.3 推定結果と精度

本論文の提案手法を用いて, 口語訳以外の新改訳と新共同訳においても原文であるギリシア語の Nestle-Aland との語彙の対応関係を抽出した. 各翻訳での精度と再現率をそれぞれ表 13, 14,15,16 に示す. また, これらの結果に基づき算出した抽出結果の正答の推定値を表 17,18 に示す. 他と同様に精度・再現率は, ランダム抽出した 100 組の単語の対応を人手で確認して算出した.

表 13 漸近推定による精度(新改訳)

相互情報量	語数精度 Pn		頻度精度 Pf	
	全体	部分	全体	部分
0.64～	90.0%	100.0%	98.5%	100.0%
0.32～	95.2%	100.0%	99.3%	100.0%
0.16～	89.7%	100.0%	95.5%	100.0%
0.08～	91.4%	98.3%	95.5%	99.6%
0.04～	91.9%	97.3%	95.6%	99.4%
0.02～	90.0%	97.5%	95.2%	99.4%
0.01～	84.9%	91.4%	94.3%	98.2%
0.005～	81.0%	88.0%	93.6%	97.4%

表 14 漸近推定による精度(新共同訳)

相互情報量	語数精度 Pn		頻度精度 Pf	
	全体	部分	全体	部分
0.64～	87.5%	87.5%	95.2%	95.2%
0.32～	90.5%	90.5%	98.0%	98.0%
0.16～	94.7%	94.7%	98.7%	98.7%
0.08～	91.5%	91.5%	97.4%	97.4%
0.04～	85.7%	87.1%	95.2%	95.6%
0.02～	81.4%	83.7%	93.5%	94.2%
0.01～	80.0%	84.2%	89.5%	94.4%
0.005～	78.0%	82.0%	88.7%	93.5%

表 15 漸近推定による再現率(新改訳)

相互情報量	頻度再現率 Rf	
	全体	部分
0.64～	19.0%	19.3%
0.32～	40.0%	40.3%
0.16～	53.4%	55.9%
0.08～	65.8%	68.7%
0.04～	71.9%	74.8%
0.02～	75.8%	79.2%
0.01～	78.3%	81.5%
0.005～	79.8%	83.1%

表 16 漸近推定による再現率(新共同訳)

相互情報量	頻度再現率 Rf	
	全体	部分
0.64～	17.4%	17.4%
0.32～	37.7%	37.7%
0.16～	52.2%	52.2%
0.08～	62.9%	62.9%
0.04～	67.9%	68.2%
0.02～	71.2%	71.8%
0.01～	72.1%	76.0%
0.005～	73.6%	77.6%

表 17 漸近推定による推定結果(新改訳)

相互情報量	累積語数精度 An		累積頻度精度 Af	
	全体	部分	全体	部分
0.64～	471	523	14,590	14,813
0.32～	1,199	1,259	30,684	30,904
0.16～	1,904	2,122	40,958	42,869
0.08～	2,824	3,037	50,485	52,656
0.04～	3,494	3,699	55,146	57,360
0.02～	3,904	4,230	58,126	60,719
0.01～	4,061	4,370	60,060	62,526
0.005～	4,174	4,535	61,201	63,694

表13から18の結果より，漸近的対応語彙推定法により，閾値 0.005 以上では 73～83%の再現率を実現していることが分かる．

頻度の精度は新共同訳では他よりも低くなるが，同程度の精度の口語訳の最尤推定の場合と比較すると再現率は 10%程度上昇している．

以降の節では本節で得られた各翻訳と原文の語彙の対応関係を分析に用いる．

表 18 漸近推定による推定結果(新共同訳)

相互情報量	累積語数精度 An		累積頻度精度 Af	
	全体	部分	全体	部分
0.64～	445	445	13,309	13,309
0.32～	1,099	1,099	28,868	28,868
0.16～	1,968	1,968	40,044	40,044
0.08～	2,757	2,757	48,254	48,254
0.04～	3,240	3,294	52,040	52,279
0.02～	3,542	3,644	54,565	55,013
0.01～	3,863	4,067	55,255	58,271
0.005～	4,076	4,285	56,412	59,467

4. 推定語彙に基づく翻訳比較

4.1 対応語彙に基づく翻訳比較

本節では，推定された原文と翻訳文の語彙の対応関係に基づき，計量的に翻訳文を比較する手法を検討する．

語彙レベルでの概念関係の翻訳において，原文の単語がどの程度忠実に，あるいは逐語的に翻訳されているかは翻訳の指針を知る上で注目すべき点である．また逆に，ある単語がどのように複数の翻訳語に訳し分けられ，あるいは原文の他の単語と同じ翻訳語に統合されるかということもまた着目すべき問題であると考えられる．

まず，推定において高い相互情報量を持つ単語の組に着目すると，これらは比較的逐語的に翻訳された単語であり，原文の意図を忠実に再現しようと強い注意が払われた単語であると考えられる．これらの単語は翻訳段階で重視された可能性がある．

また，原文のある単語と別の単語を翻訳文では同じ単語に訳す場合，翻訳者はこれらの原文の単語間に概念的類似性を感じていることになる．逆に，原文で一つの単語を翻訳文で複数に訳し分ける場合，翻訳者はこれら翻訳語に関連を見出しつつも区別の必要性を感じていると言えよう．このため本論文で先に抽出した原文と翻訳文での語彙の対応関係のデータは，翻訳において前提とされた概念の関係性を反映していると考えら

れる. 多数の単語の対応組が総体として構成する複雑な概念関係構造を分析することで, 翻訳における背景的な理解の差異を抽出できる可能性がある.

以下では, これらの方針に基づき, 相互情報量の高い単語, 原文と翻訳文での語彙の対応関係の分析を行う.

4.2 原文との高一致度単語

まず原文と翻訳文の対応する語彙間で相互情報量の多い単語, すなわちギリシア語の語義のカバーする範囲が忠実に日本語に翻訳されている単語を抽出する. これらの単語は翻訳上, 逐語訳を強く意識された単語であると考えられる.

表 19 相互情報量の高い共通単語

原語	頻度	日本語訳(口語訳)
ἀδελφός	343	兄弟
προφήτης	144	預言者
ἀγάπη	116	愛
ἐκκλησία	114	教会
αἷμα	97	血
σάββατον	68	安息日
γραμματεύς	63	律法学者
ποτήριον	31	杯
συνείδησις	30	良心
κώμη	27	村
νεφέλη	25	雲
σωτήρ	24	救主
ἀμπελών	23	ぶどう園
τελώνης	21	取税人
ἵππος	17	馬
νυμφίος	16	花婿
διψάω	16	かわく

日本語の三つの翻訳文全てにおいて, 相互情報量が 0.8 以上でかつ, 単語の頻度が 15 以上で固有名詞を除いたものを表 19 に示す. 表 19 より頻度の高い単語(30 以上)は全てキリスト教用語であると分かる. これらの単語は, 宗教概念を翻訳する上で重要なため, 訳語の共通化が意識的に図られたと考えられる.

次に, 一つの翻訳では相互情報量が高いが他では低い単語を抽出する. このような単語は, 各翻訳で重視される概念の相違を反映すると期待される. 具体的には, 原文で頻度 15 以上の単語の中で, 一つの翻訳での相互情報量が 0.2 以上かつ他の二つの翻訳の相互情報量より 1.5 倍以上大きい単語を抽出する. 抽出結果をそれぞれ表 20, 21, 22 に示す.

表 20 口語訳のみ相互情報量が高い単語

原語	頻度	口語訳	相互情報量
περιπατέω	95	歩く	0.603
θάλασσα	91	海	0.865
θλῖψις	45	患難	0.512
εὐχαριστέω	38	感謝する	0.620
μέλος	34	肢体	0.882
παρουσία	24	来臨	0.458
ἐπίγνωσις	20	知識	0.417
παράπτωμα	19	罪過	0.607
σκοτία	16	やみ	0.630
οἰκουμένη	15	世界	0.533

表 21 新改訳のみ相互情報量が高い単語

原語	頻度	新改訳	相互情報量
γεννάω	97	生まれる	0.591
γαμέω	28	結婚する	0.526
θερίζω	21	刈り取る	0.667
ἄφεσις	17	赦す	0.401

表 22 新共同訳のみ相互情報量が高い単語

原語	頻度	新共同訳	相互情報量
πνεῦμα	379	霊	0.558
εἰσέρχομαι	194	入る	0.438
κόσμος	186	世	0.513
εἰρήνη	92	平和	0.722
κάθημαι	91	座る	0.408
κηρύσσω	61	宣べ伝える	0.483
ἐπιθυμία	38	欲望	0.406
ὁμολογέω	26	言い表す	0.505
ἰάομαι	26	いやす	0.572
τράπεζα	15	食卓	0.467
ἀνέχω	15	我慢する	0.474

表 20 中でキリスト教用語として頻繁に用いられるのは「感謝」「罪過」と「来臨」であるが, 「感謝」「罪過」を意味する単語は多数あるため, これらを同じ単語に訳す他の翻訳と, 特定の組を強く結び合わせた口語訳で差が現

れたと考えられる.「来臨」はキリストが再び到来する信仰を表す特徴的宗教用語であり,類似単語と混同することは考えにくい. 新改訳では「来臨」は使うが口語訳に比べその頻度が少なく(5 回), すなわち「来臨」の概念が示す範囲が口語訳に比べて狭いことが分かる. しかし新共同訳では「来臨」の語は使わずに単に「来る」と訳しており, 訳語選択から「来臨」を何かしら特別な宗教用語で明示的に表現する必要性が, 他の翻訳に比べるとあまり認識されてはいないと言えよう. 一般的に, 注意が強く払われている概念ほど, より詳細に多数の単語を用いて表現し分けられる傾向が知られており[21], 他の二つの翻訳が単語レベルで明示的に単に「来る」ことと「来臨」を峻別しているのに比較すると, 新共同訳では「来臨」信仰が重視されていないと推測される.

表 21 中での頻出のキリスト教用語は「刈り取る」と「赦す」であるが,「刈り取る」は主に終末の審判での罪の裁きに関係して用いられる用語である.「赦す」も罪に関係した単語であり, 新改訳は他翻訳よりも罪と赦し・裁きに注意を払った翻訳であると考えられよう.

表 22 中では「霊」「平和」「宣べ伝える」「言い表す」などはキリスト教用語であるが,「宣べ伝える」「言い表す」は信仰の告白と布教を意味する単語であるため, 宣教への意識が表れていると言えよう. また,「霊」と「平和」は,「霊」により交わり「平和」を保つエキュメニズム思想[20]と関係があり, エキュメニズム思想を背景として成立した新共同訳の特色が現れていると考えられる.

4.3 対応語彙ネットワーク

4.3.1 ネットワークの作成

次に, 原文での一つの単語が翻訳では複数単語に訳し分けられる場合とその逆の場合の対応関係を分析する. 原文と翻訳文の語彙の対応関係は複雑な多対多の対応関係であるが, このような人文学的な複雑な概念間の関係性の分析にはネットワーク分析が有効である[22,23]. 意味的関連の強い単語の結合でできるネットワークは, 認知科学や人工知能学等の分野で用いられる, 概念関係を示す意味ネットワーク／概念グラフ[24]である. 意味ネットワークを作成することで, 種々のネットワーク分析技法を用い, 計量的に概念関係を分析可能となる.

本論文で扱う原語と翻訳語の対応関係は bipartite graph になっているため, 原文の語彙ネットワークと翻訳文の語彙ネットワークの二つを作成可能である. 以下ではこれらの二つのネットワークを各翻訳から作成して, 翻訳間の比較を行う.

二種類のネットワーク作成の経過を図 6, 7, 8 に示す. まず図 6 が示すのは相互情報量に基づいて抽出された単語間の対応関係である. 次に, 翻訳語の単語によって結ばれる原語の単語をネットワーク化したのが図 7 である. 図の点線円は翻訳語の単語によって結合された原語の単語の範囲を示している. 逆に原語の単語により結合されたネットワークが図 8 となる. ネットワーク中の単語 a,b 間のエッジのウェイト Wab は, a と b を仲介する他言語の単語 i と a と b の相互情報量をそれぞれ P(a,i), P(b,i)とし, 仲介する語数を n とすると,

$$W_{ab} = n \times \sum_{i} P(a,i) \times P(b,i)$$

と表現できる. この式は, a と b が均等に n 個の単語を経由する場合も, 1 個の単語のみを経由する場合にも同じウェイトを与える. 単語の対応関係は 3.3 節の分析で得られた対応の中で相互情報量 0.02 以上の単語の対応関係から誤った対応関係を人手で除いたものを用いた.

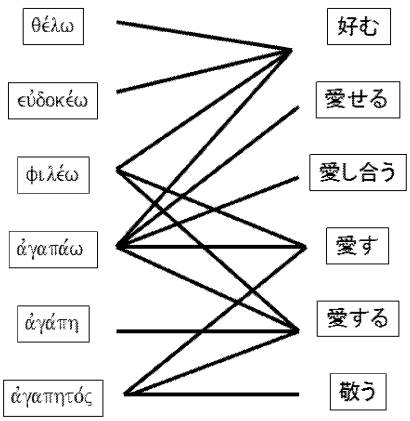


図 6 語彙間の対応例

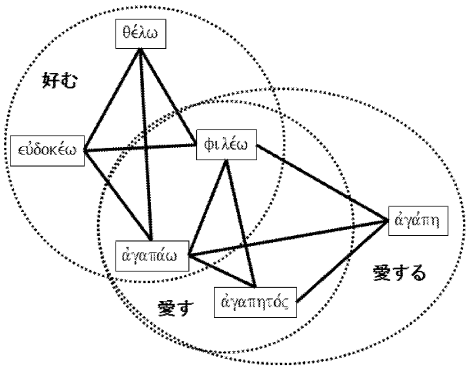


図 7 原文の語彙ネットワーク例

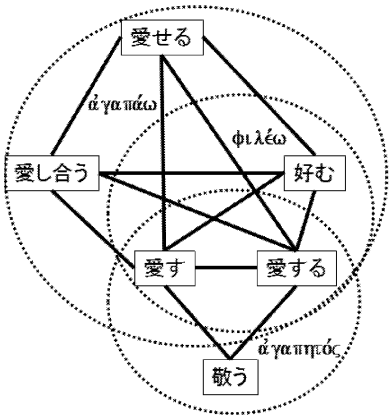


図 8 翻訳文の語彙ネットワーク例

表 23 ネットワークの基本情報

	翻訳	対象	ノード	エッジ	密度
原文	口語訳	全体	395	595	0.0076
		連結	230	437	0.0166
	新改訳	全体	397	607	0.0077
		連結	217	404	0.0172
	新共同訳	全体	409	655	0.0079
		連結	290	565	0.0135
翻訳	口語訳	全体	1,188	2,291	0.0032
		連結	507	1,378	0.0107
	新改訳	全体	1,151	2,185	0.0033
		連結	437	1,119	0.0117
	新共同訳	全体	1,193	2,361	0.0033
		連結	609	1,695	0.0092

結果として得られる6個のネットワークの基本統計量は表 23 に示す. 表 23 では, 作られたネットワークのデータを「全体」の項目に, ネットワーク中の最大連結部分グラフのデータを「連結」の項目に示した.

得られるネットワークの例として, 口語訳と新改訳の訳語から, 一部分を比較のため抜粋したものを図 9, 10 に示す.

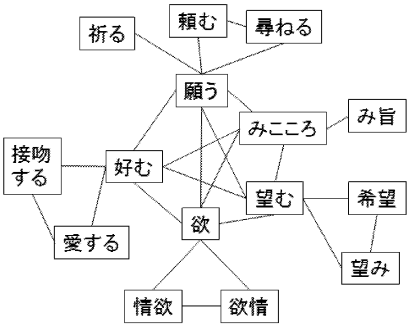


図 9 口語訳の語彙ネットワーク部分図

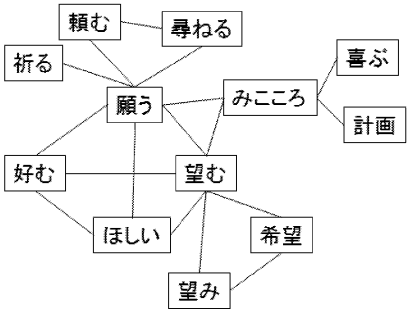


図 10 新改訳の語彙ネットワーク部分図

図 9, 10 より, 同じ「願う」「望む」などに関する概念は, 非常に類似した概念の関係性を含むことが分かる. 一方で口語訳では「愛」や「情欲」にも結び付くなど, 異なる部分も少なくない. これらの相違は, 各翻訳での概念のニュアンスの差異を示唆するデータであり, 各概念の微細な意味関係の差異が全体としての概念関係のネットワーク構造にも差異をもたらしている. これらの構造の差異は, 概念関係の構造の差異, すなわち原文理解での意味構造の各翻訳での差異を反映していると考えられる.

4.3.2 ネットワークの中心性

ネットワークを作成したのみでは, 人が見て大まかな特徴と相違を発見することは可能であるものの, それは数値的正確さや客観性に欠ける. そこで原文と翻訳文の対応関

係から得られた2つのネットワークの特徴を計量的に比較するため，ネットワークの分析では一般的な指標である，中心性の計算と比較を行う．

ネットワークにおける中心性は，多数の要素が含まれる複雑な関係性の中で，数値的に各要素の「重要度」を計算する目的で使われている．社会の人間関係をネットワーク分析する分野では，多くの人物とコネクションを持つという意味での重要性（Degree），どの人物からも関係性が相対的に近いという意味での中心性（Closeness），人と人とのパイプ役としての貢献の大きさ（Betweeness），社会的に重要な人物と太いパイプを持つという意味での重要度（Bonacich Power）などさまざまな中心性の計算方法が考案され，分析に用いられている．これらの中心性の計算はネットワーク構造を持つ対象であれば同様に実施可能であるが，ネットワークが何を表現したかによってこれらの中心性の示す意味は異なってくる．

翻訳語彙の対応関係から導出された，概念的類似関係にある単語を結合したネットワークでの中心性とは，すなわち，複雑な概念の相関関係において，翻訳者が何らかの意味で重視をしている要素であると考えられる．どのような意味での重要性が抽出されるかは，各中心性のアルゴリズムに依存する．本論文では，ネットワークの中心性分析で一般的に用いられる，Degree, Closeness, Betweeness, Bonacich Power の四種類を分析に用いる[25,26]．

これらの中心性の翻訳対応語彙関係ネットワーク上での意味としては，Degree では多数の単語と対応する概念範囲の広い単語が上位に来ることが予想され，概念範囲の広さの指標になりうると予測される．Closeness では，全概念関係の中央に位置する単語が抽出されやすいため，翻訳者の世界観の中心的な単語が抽出される可能性がある．Betweeness では，全概念関係を接続させるハブ的な役割を果たす単語の抽出が期待できる．Bonacich は相互の影響力が加味された形で中心的単語の抽出が行われると期待される．中心性は UCINET[27]を用い，最大連結部分グラフに対して計算した．各中心性の結果上位 10 位までを表 24, 25, 26, 27 に示す．

表 24 より，Degree では「行く(ἔρχομαι)」「知る(οἶδα)」「見る(ὁράω)」「言う(λέγω)」「分かる(οἶδα)」「願う(θέλω)」などの一般的内容を示す動詞や形容詞の中心性が高くなることが分かる．これらの約し分けにより多くのノードが他と結合されている状況がうかがわれる．

表 25 より，Closeness では各翻訳での特徴的単語が出ている．まず口語訳は「尋ねる」「求める」「ねらう」「捜す」「要求する」などの何かを得ようと探し求める単語が多く，「努める」「かなえる」も願いとその実現に関連した単語と考えられる．ギリシア語でも ζητέω, αἰτέω , ἐρωτάω , ἐπερωτάω , θέλω , βούλομαι が尋ね願い求めることに関した単語である．

新改訳では，「知らせる」「伝える」ことに合わせて，「知る」「分かる」「認める」「ご存じ」「確かめる」「気づく」「悟る」等の理解することに関する単語が上位に来ている．これは教えが伝えられ理解されることの重視を示すと考えられる．原語でも，伝えること（ἀναγγέλλω , ἀπαγγέλλω , κηρύσσω , παραδίδωμι）と知り理解すること（γινώσκω, ἐπιγινώσκω, γνωρίζω）に関した単語が上位を占める．

新共同訳でも，新改訳と同じく「知る」や「認める」も含みつつ，「考える（λογίζομαι）」も上位に来ている．新改訳も新共同訳も単に「知り」「認める」のではなく「考え」「悟る」ことが重視されていると考察される．これは中世神学よりの伝統である「知解を求める信仰」[28]の概念に適合する．また，「先（ἀρχη,）」，「最初（πρῶτος）」や「支配（ἄρχων）」が上位である点も特徴的である．最初の存在である神の支配を強調しているのであろうか．

表 26 より，Betweeness でも各翻訳に特徴的な単語が見られる．口語訳では，Closeness と同様に「求める」ことに関する単語が多いが，それ以外に「集める」や「集まる（συνάγω）」，「呼ぶ」などの単語が見られる．これらはいずれも，呼ばれ集まる人々，すなわち教会（ἐκκλησία）を示唆する単語である[29]．口語訳では神に呼ばれ集まる者としての教会が幾多の神学的概念を結びつけるハブ的な役割を果たしていると考えられる．

表 24 Degree 中心性の比較

口語訳				新改訳				新共同訳			
行く	4.94	θέλω	6.99	知る	5.73	ὁράω	8.80	行く	4.28	ὁράω	6.23
まさる	4.35	λέγω	6.11	捨てる	5.28	γινώσκω	6.48	知る	4.28	γινώσκω	5.19
尋ねる	3.76	ἔρχομαι	5.24	わかる	5.05	ἐπιγινώσκω	5.09	認める	3.62	βλέπω	5.19
呼ぶ	3.56	ζητέω	5.24	取る	4.59	ἀρνέομαι	5.09	殺す	3.62	ἀγνοέω	4.15
願う	3.56	γινώσκω	4.37	願う	4.13	βούλομαι	5.09	立ち去る	3.45	ἄγω	4.15
見る	3.36	ἀπέρχομαι	3.93	上げる	3.90	θέλω	5.09	現れる	3.29	αἴρω	4.15
聞く	3.36	ὑπάγω	3.93	いっぱい	3.90	ἔρχομαι	5.09	取る	3.13	ἐπιγινώσκω	4.15
思う	3.36	προσέρχομαι	3.93	知らせる	3.44	οἶδα	4.63	全	2.96	οἶδα	4.15
帰る	3.36	βούλομαι	3.93	来る	3.44	βλέπω	4.63	全体	2.96	ἐξέρχομαι	3.81
大いなる	3.16	φρονέω	3.93	頼む	3.44	αἴρω	4.63	だれ	2.80	ἐπέρχομαι	3.81
現れる	3.16	παραλαμβάνω	3.93	帰る	3.44			どんな	2.80	ἔρχομαι	3.81
話す	3.16			現われる	3.44						
				取りのける	3.44						

表 25 Closeness 中心性の比較

口語訳				新改訳				新共同訳			
尋ねる	18.29	ζητέω	19.00	知らせる	17.16	γινώσκω	16.76	考える	16.70	λογίζομαι	17.83
求める	18.28	αἰτέω	18.54	知る	16.47	ἀναγγέλλω	16.71	認める	16.68	πρῶτος	17.48
引き取る	17.86	αἴρω	18.07	わかる	16.28	ἀπαγγέλλω	16.63	おもだった	16.31	γινώσκω	17.40
思う	17.58	ἐρωτάω	17.88	伝える	16.25	παραδίδωμι	15.97	支配	16.24	βλέπω	17.39
集める	17.19	λέγω	17.81	御霊	16.12	ἐπιγινώσκω	15.85	先	16.21	ἀγνοέω	17.32
努める	16.95	ἐπερωτάω	17.14	ゆだねる	15.78	κηρύσσω	15.84	最初	16.11	δοκέω	17.29
ねらう	16.92	θέλω	17.10	認める	15.62	πνεῦμα	15.82	知る	16.11	ἄρχων	17.20
計る	16.92	ἡγέομαι	17.09	ご存じ	15.48	γνωρίζω	15.81	見なす	16.08	ἡγέομαι	17.16
捜す	16.92	συνάγω	16.88	確かめる	15.48	πιστεύω	15.71	終わる	15.97	ἀρχή	17.07
かなえる	16.58	βούλομαι	16.72	気づく	15.48	λογίζομαι	15.51	行く	15.82	μαρτυρέω	16.77
要求する	16.58	φρονέω	16.72	悟る	15.48						

表 26 Betweeness 中心性の比較

口語訳				新改訳				新共同訳			
求める	49.35	ζητέω	54.16	知らせる	53.84	πνεῦμα	47.18	認める	31.13	βασιλεία	25.03
引き取る	49.19	αἴρω	51.49	御霊	43.10	ἀναγγέλλω	44.96	支配	29.66	ἔθνος	24.50
尋ねる	44.72	αἰτέω	49.30	たましい	37.59	ψυχή	40.85	国	25.50	λογίζομαι	22.12
集める	40.72	λέγω	30.17	いのち	35.68	ζάω	37.74	同胞	21.95	μαρτυρέω	20.52
集まる	28.23	συνάγω	28.21	生き返る	32.02	πείθω	34.83	評判	17.82	γένος	19.49
着く	23.35	παραγίνομαι	28.00	ゆだねる	29.96	ἐγείρω	33.77	いろいろ	17.59	ἄρχων	19.33
思う	19.24	ἔρχομαι	22.89	立つ	29.53	παρίστημι	31.47	考える	17.47	ἀκοή	18.12
頼む	16.91	ἐρωτάω	19.06	伝える	29.28	γινώσκω	31.45	行く	17.27	ἀρχή	17.70
話す	15.29	πιστεύω	18.53	ささげる	28.85	προσφέρω	31.24	終わる	17.09	ἀκούω	17.63
呼ぶ	14.91	διακονέω	15.30	連れる	28.54	παραδίδωμι	28.97	連れる	15.96	πείθω	16.43

表 27 Bonacich 中心性の比較

口語訳				新改訳				新共同訳			
多く	4.94	οἶδα	2.09	ひとり	1.01	οἶδα	2.16	すべて	9.47	οἶδα	2.26
多数	4.80	γινώσκω	2.04	たたむ	1.01	γινώσκω	2.09	何もかも	9.28	γινώσκω	2.05
大ぜい	4.35	ἀγνοέω	1.60	持ち去る	1.01	ὁράω	1.53	あらゆる	9.19	ἐπιγινώσκω	1.70
たくさん	4.00	ἐπίσταμαι	1.53	同体	1.01	θεωρέω	1.47	皆	8.80	μαρτύριον	1.58
大いなる	3.97	ὁράω	1.14	聖徒	1.01	ἀγνοέω	1.32	みんな	8.44	μαρτυρέω	1.53
見いだす	3.54	ἐπιγινώσκω	1.12	聖霊	1.01	ἐπίγνωσις	1.32	万物	8.44	δουλεύω	1.52
見つかる	3.54	λέγω	1.12	一瞬	1.01	οἶκος	1.27	みな	8.11	διάκονος	1.43
見当る	3.54	παιδίον	1.09	見いだす	1.01	οἰκία	1.27	全部	7.74	χαίρω	1.42
きたる	3.52	φημί	1.08	見つかる	1.01	ἐπιγινώσκω	1.24	めいめい	7.04	οἶκος	1.32
多い	3.42	βλέπω	1.01	見つける	1.01	βλέπω	1.22	一切	6.56	οἰκία	1.31

新改訳では，先の Closeness の結果に合わせて「たましい（ψυχή）」「いのち（ζάω）」「生き返る（ἐγείρω）」等の単語が上位に来ており，また，「御霊（πνεῦμα）」も新たに含まれている．「御霊」は，それが注ぎ込まれることで人間が命を得ると考えられており[30]，人間は神によって生かされ，復活させられることが重要な概念であると言えよう．これは新改訳の背景となっている福音主義的思想での超自然的力を強調する傾向と合致する．

新共同訳では「国（βασιλεία, ἔθνος）」「支配（ἄρχων）」が上位に来ており，また「評判（ἀκοή）」や他にも聞くことに関する単語（ἀκούω，πείθω）も上位に含まれる．これらは「神の国・支配の到来の知らせ」すなわち「福音」[31]が他の神学的概念を結びつける役割を果たしていることを示唆すると考えられる．

Bonacich では，ギリシア語では οἶδα，γινώσκω，ἐπιγινώσκω（「知る」）が全翻訳で上位に含まれている．口語訳と新改訳では ὁράω と βλέπω（「見る」）が共通であり，οἶκος と οἰκία（「家」）は新改訳と新共同訳に共通する．これより「知る」ことが新約聖書全体の重要事項であると言えよう．

各日本語では特別な宗教用語ではない多寡を表す表現が多く，恐らく各翻訳の表現の特徴を表すものと推測される．新改訳のみ聖なるものに関する単語が複数含まれているが，神の超自然的力を強調する福音主義的思想により聖なるものに関する表現が重視されている結果と推測できる．

これらの結果より，三翻訳の共通的な関心事項は「知る」ことにあると言えよう．また，各翻訳の個別的な特徴としては，まず口語訳では，「尋ね探し求める」ことと「呼ばれ集められたものとしての教会」が中心的な課題であると考えられる．

新改訳は，「伝え知らせる」ことが重視されると同時にそれらを「確かめて悟る」ことが意識されている．また命の源としての神とその霊や聖なる存在が強調される傾向が強い．

新共同訳では，「考えて認める」ことが重視されている．また「神の国の到来に耳を傾ける」ことが強調されていると言えよう．

以上より得られた各翻訳の特徴を表 28 に示す．

表 28 比較結果のまとめ

	口語訳	新改訳	新共同訳
高一致単語	来臨	罪の赦しと裁き	霊による平和、宣教
Closeness	探し求める	宣教と知解	知解
Betweeness	呼ばれた者としての教会	霊によって与えられる命	神の国・支配の到来
Bonacich	知ること		

5. 結論と今後の展望

本論文の成果は，大きく以下の三点に分けられる．

1. 原文と翻訳文の対応語彙の特定・抽出アルゴリズムである漸近的対応語彙推定法を考案した．考案手法は，分析の対象テキスト以外に他の大規模テキストコーパスや学習用データの利用が困難なテキストにおいても利用が可能であり，かつ従来の手法に比べて再現率が 20%程度向上している．また，最尤推定法等と異なり，一つの単語に対する一つのみの翻訳語を特定するのではなく，複数の単語に訳し分けられる場合にもそれらの対応関係を適切に抽出することが可能である．漸近的対応語彙推定法は単語の一対一対応の仮定，被特定単語の除去による相互情報量の再計算，閾値の漸近的低減を組み合わせることでこれらの性能実現した．

また本論文では漸近的対応語彙推定法の適用により，聖書の日本語翻訳より頻度精度で 95%以上，頻度再現率で 78%以上（相互情報量閾値 0.01 の場合）の原語と翻訳語の対応語彙の推定が実現できた．

2. 推定された語彙の対応関係を利用して複数の翻訳文間の語彙レベルでの計量的比較を行った．原文から逐語的に翻訳されている単語，すなわち注意を払って原文に忠実に翻訳された単語の抽出により，翻訳で重視された特徴的な思想の差異を分析した．結果として，口語訳は再臨の強調，新改訳は罪の赦しと裁き，新共同訳は霊による平和な一致が個別的特徴であることが示唆された．

3. 各翻訳での語彙の対応関係から，背景的な概念間の構造を，原語と翻訳語の 2 種類のネットワークとして表現した．また，ネ

ットワークに対して四種類の中心性の分析を行い，概念の関係性の中で中心的な位置を占める単語を抽出した．

結果として，三翻訳共通に「知る」ことが重視されているが，「探し求めて」「知る」か（口語訳），「伝えられて」「知る」か（新改訳），「知り」「考える」（新共同訳）のかというように，重視される内容は近くともそのニュアンスが異なることが明らかとなった．

他にも口語訳は「呼び集められたもの」，新改訳は「霊によって与えられた命」，新共同訳は「神の国の到来」が概念の関係の中で重要な位置を占めていることが明らかとなった．

抽出された概念的な特徴は，これらの翻訳作成の背景的な思想（福音主義，エキュメニズム）とも合致している．よって，原文と翻訳文の語彙の対応関係の相違より翻訳の背景にある思想的な特徴の相違を計量的に抽出することは可能であると結論できる．

本研究の今後の課題としては以下の点が挙げられる．

1．本論文においては，語彙の対応関係を一対一に限定したが，共起語の辞書やフレーズ辞書などを構築し，複数単語が一単語と対応する場合も的確に抽出できるようにアルゴリズムを改変して精度と再現率をさらに向上させる．

2．同じ単語の名詞形と動詞形などを（「愛」と「愛する」など）グルーピングすることで，文中での品詞に関わらず概念間の関係性を精度よく抽出できる可能性がある．

3．得られたネットワークに対してエゴセントリックネットワーク分析を用いることで特定単語の解釈のニュアンスの差異の計量的な分析や，クラスタリング手法の適用による概念のグループ化が可能と考えられる．

本研究で提示した手法は，聖書以外でも，相異なる解釈に基づく複数の翻訳が並立する，多くの古典的テキストや，著名な文学・思想関連のテキスト群にも適用が可能である．本手法の適用により従来人文学的手法でしか行えなかった分析を，より客観的かつ精密・大規模に実現する道が開かれると期待される．

謝辞

本研究は科研費「レトリカルデータベースシステムの構築による計量的修辞分析手法の確立」（22700256）の助成を受けた．

参考文献

[1] 矢崎祐一:英日翻訳における聖書引用箇所の訳文—詩篇 23 篇の場合—，翻訳研究への招待, No. 4, pp. 107-122, 2010.
[2] Brown, Peter; Della Pietra, Vincent; Della Pietra, Stephen; Mercer, Robert: “The Mathematics of Statistical Machine Translation: Parameter Estimation”, Computational Linguistics, Vol. 19, No. 2, pp.263‐311, 1993.
[3] Koehn, Philipp: “Europarl: A parallel corpus for evaluation of machine translation”, In Proceedings of MT Summit, 2005.
[4] 宇津呂武仁; 日野浩平; 堀内貴司; 中川聖一: “日英関連報道記事を用いた訳語対応推定”, 自然言語処理, Vol. 12, No, 5. pp. 43-69, 2005.
[5] Rapp, Reinhard: “Automatic Identification of Word Translations from Unrelated English and German Corpora”, Proceeding of 37th ACL, pp.519-526, 1999.
[6] 森下洋平; 宇津呂武仁; 山本幹雄: “対訳特許文書からの専門用語対訳辞書半自動獲得におけるフレーズテーブルと既存対訳辞書の併用”, IPSJ SIG Notes No. 90, pp. 91-98, 2008.
[7] Papineni, Kishore; Roukos, Salim; Ward , Todd; Zhu, Wei-Jing: “BLEU: a method for automatic evaluation of machine translation”, Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pp. 311-318, 2002.
[8] Culy, Christopher; Riehemann, Susanne: “The Limits of N-Gram Translation Evaluation Metrics”, In Proceedings of the Ninth Machine Translation Summit, 2003.
[9] 日本聖書協会: “口語訳聖書”, 日本聖書協会, 1955.
[10] 日本聖書刊行会: “新改訳聖書”, 日本聖書刊行会, 1970.
[11] 日本聖書協会: “新共同訳聖書”, 日本聖書協会, 1987.
[12] 鈴木範久, “聖書の日本語”, 岩波書

店, 2006.
[13] 新改訳聖書刊行会: “聖書翻訳を考える”, いのちのことば社, 2004.
[14] 日本コンピュータ聖書研究会: “JBible”, http://www.jbible.net/jcbr/, 2010/6/23 参照.
[15] Nestle, E.; Aland, B: “Nestle-Aland Novum Testamentum Graece”, 1982.
[16] BibleWorks, http://www.bibleworks.com/, 2010/6/17 参照.
[17] 川島隆徳; 村井源: TextSeer, http://www.valdes.titech.ac.jp/~t_kawa/, 2010/6/17 参照.
[18] Gale, William; Church, Kenneth: “Identifying Word Correspondences in Parallel Texts”, Proceedings of the 4th DARPA Speech and Natural Language Workshop, pp. 152–157, 1991.
[19] Gale, William; Church, Kenneth; Yarowsky, David: “Using bilingual materials to develop word sense disambiguation methods”, Proceedings of the International Conference on Theoretical and Methodological Issues in Machine Translation, pp. 101-112, 1992.
[20] パウロ 6 世;”エキュメニズムに関する教令”, 第 2 バチカン公会議文書, 1964.
[21] Sapir, E.; Whorf, B. L.(池上嘉彦訳): 「文化人類学と言語学」, 弘文堂, 1995.
[22] 村井源; 徃住彰文: "正典テキスト群から編集的中心メッセージを抽出するネットワーク解析法", 情報知識学会誌, Vol. 17, No. 3, pp. 149-163, 2007.
[23] 村井源; 川島隆徳; 徃住彰文: "医療の質・安全研究における関心領域の分析-学術論文の計量的分析による研究動向の抽出-", 医療の質・安全学会誌, Vol. 4, No. 1, pp.16-24, 2009.
[24] Sowa, John F.: “Conceptual graphs as a universal knowledge representation”, Computers & Mathematics with Applications, Vol. 23, Issues 2-5, pp. 75-93, 1992.
[25] 村井源;山本竜大; 徃住彰文: "Web サイトデータを活用した計量的人間関係解析のための指針—日本の国会議員 Web サイトからみた政治家の中心性とグループ—", 理論と方法, vol. 23, no. 1, pp. 110-128, 2008.
[26] Bonnacich, P: “Power and Centrality: A family of Measures”. American Journal of Sociology No. 92, pp. 1170-1182, 1987.
[27] Analytic Technologies; “UCINET”, http://www.analytictech.com/ucinet/, 2010/6/25 参照.
[28] リーゼンフーバー, K: “知解を求める信仰”, ドン・ボスコ社,2005.
[29] 新約聖書; “マタイによる福音”, 13 章 2 節など.
[30] 旧約聖書; “創世記”, 2 章 7 節など.
[31] 新約聖書; “マタイによる福音”, 4 章 23 節など.

(2010年6月28日受付)
(2010年8月13日採択)
(2010年8月26日早期公開)

次世代情報処理基盤としての リファレンス構造システムとマイクロデータ
Reference Structure Systems and Micro-data as Next Generation Information Infrastructure

渡邊修[1*], 高島史郎[2], 今津達也[3]
Osamu WATANABE, Shiroh TAKASHIMA, Tatsuya IMAZU

*1 ディスクロージャー・イノベーション株式会社 マイクロデータ研究会
Disclosure Innovation Inc. , MicroData Consortium
〒171-0033 東京都豊島区高田 3-28-8
E-mail：owatanabe@di-inc.co.jp
2 株式会社 KDC マイクロデータ研究会
KDC Inc. , MicroData Consortium
〒202-0003 東京都西東京市北町 5-2-9
E-mail：QZD06271@ nifty.com
3 ゼッタテクノロジー株式会社 マイクロデータ研究会
Zetta Technology Inc. , MicroData Consortium
〒113-0022 東京都文京区千駄木 3-47-1 千駄木 WIN ビル
E-mail：timazu@zetta.co.jp

これまで，標準化によるデータ共有のための試行錯誤が繰り返されてきたが，程度の差はあるにせよ，データを生成場所から特定のデータベースシステムに移動(＝コピー)して集中させる発想が根底にあった．同じ意味のデータが複数存在することは，データ間の同期が保証される仕組みが確立されていれば問題が生じることはないが，多くの場合，同期が確保されているとは言い難い状況である．

リファレンス構造システムとマイクロデータは，これまでのデータ標準化と共有化のための取り組みの中で克服することができないままでいる上記課題を解消するために構想されたものである．

Data sharing realized through standardization has so far been tried, and to some extent, this is based on the idea of copying the data at the origin and then moving it to specific database systems. In this system, copies of data having the same meaning exist in several locations. It causes no problem so long as the synchronization of the data is guaranteed. However, in most cases such synchronization is not ensured.

Reference structure systems with Micro-data are designed to solve the above problem regarding data standardization and data sharing.

キーワード:リファレンス構造，マイクロデータ，標準化，データベース
Keyword：Reference structure， micro-data， standardization， database

1 はじめに

現在，我々が扱っているデータは，基本的に業務単位で構築されたコンピュータシステムで生成され記録されている．コンピュータネットワークが未発達な時代であればコンピュータシステムが業務単位で構築されたことは不可避的であったかもしれないが，コンピュータシステムが業務単位で無秩序に構築されてきた結果，データの存在そのものがコンピュータシステムに依存してしまい，業務を跨いだデータの共有

が事実上不可能になっている．

政府が実施している統計調査を例にすると，各府省の担当部署単位で調査票が設計されており，当初から調査項目が標準化され政府内部でデータが共有化されているわけではない．その結果，同じ調査客体（民間事業者や地方自治体）に対して複数の調査主体（政府の統計調査担当部署）が同じ内容の調査を実施することにもなり，法的に同じ意味の調査項目であるにも関わらず調査主体が異なれば調査結果の値が異なるという事態が生じている．統計調査以外の行政事務においてもコンピュータシステムが乱立しており，別途データの同期を担保する仕組みがなければ，データの非同期による誤判断というリスクが伴う．

このような状況を克服するためにデータの標準化がこれまで幾度となく試みられてきたが，データを標準化し業務横断的に共有化するためには業務間での調整が不可欠であり，既存システムの改修も避けて通れない．しかも，個々の業務単位ではデータが標準化され他の業務と共有化されていなくても現状の業務を実施するうえでは支障が生じないことから，データ標準化のための労力やシステム改修のコストを考慮した結果，多くの場合あえて手を付けないでいるのが現状である．

また，標準化されたデータを扱うために巨大なデータベースシステムやメッシュ型システムが構想されてきたが，システムそのものの在り方や登録されるデータの管理をめぐる議論が尽きないなどの理由により，なかなか構築できないでいる．

このように，これまで標準化によるデータ共有のための試行錯誤が繰り返されてきたが，程度の差はあるにせよ，データを生成場所から特定のデータベースシステムに移動（＝コピー）して集中させる発想が根底にあった．

同じ意味のデータが複数存在することは，データ間の同期が保証される仕組みが確立されていれば問題が生じることはないが，多くの場合，同期が確保されているとは言い難い状況である．データを生成した者が自己のデータに誤りを発見して修正しても，データのコピー先であるデータベースシステムのデータが修正されなければ，関係者は誤ったデータのまま業務を続けるという危惧がある．

2 これまでの経緯

このようなコンピュータ間においてデータ共有がなされていない状況は，データを取得する際に費やす夥しい時間や定義の差から生じるデータ精度の低下により，特に加工統計（二次統計）を生成する場面で弊害として顕在化してきている．このため，2006年度に内閣府経済社会総合研究所は「電子化に対応した経済社会統計のあり方研究会」を設け「SNAにおけるマイクロデータの活用に関する理論的検討」として検討結果を取りまとめるとともに，マイクロデータを SNA に導入する際の課題を検証するための実証実験として「電子化に対応した経済社会統計のあり方の調査研究」を実施している[1]．

こうした内閣府の調査研究が契機となり，20008 年 1 月には民間企業有志により「マイクロデータ研究会（当初は「次世代情報システム基盤研究会」）」が発足した．

その後，マイクロデータ研究会として調査研究を積み重ねた結果，2009 年 2 月 10 日に開催された高度情報通信ネットワーク社会推進戦略本部（IT 戦略本部）の第 6 回次世代電子行政サービス基盤等検討プロジェクトチームにおいて，データ標準化の取り組み例として政府に対しリファレンス構造

システムとマイクロデータの有効性を説明する機会を得ている[2].

3 次世代情報処理基盤の概要

リファレンス構造システムとマイクロデータはこれまでのデータ標準化と共有化のための取り組みの中で克服することができないままでいる課題を解消するために構想されたものであり，以下，その概要について説明する.

3.1 リファレンス構造システムとマイクロデータ

3.1.1 これまでのデータ流通について(従来型システムの問題点)

情報処理システムは“処理”をどこでどのように行うかによって，集中型，分散型等に分類することができるが，そのシステムが扱う情報/データはホスト・コンピュータやサーバに蓄積され，蓄積したものを処理することが一般的であり，分散型システムでも情報/データは特定のサーバに集められ，結合・集約・集計を行ってその結果を提供している.

集められた情報/データの加工(結合・集約・集計)とは，その情報/データの発生要因や情報/データが持つ価値や性質とは関係なしに特定の目的を達成するために処理していると考えることができる．このような現在の情報処理システムは合目的的に作られているため，情報/データの提供側や利用側のどちらかが相手のシステムに合わせなくてはならない．提供先や提供元が追加になるのであれば，時間と手間をかけて新しいシステムや機能を構築・追加することになる(図1を参照).

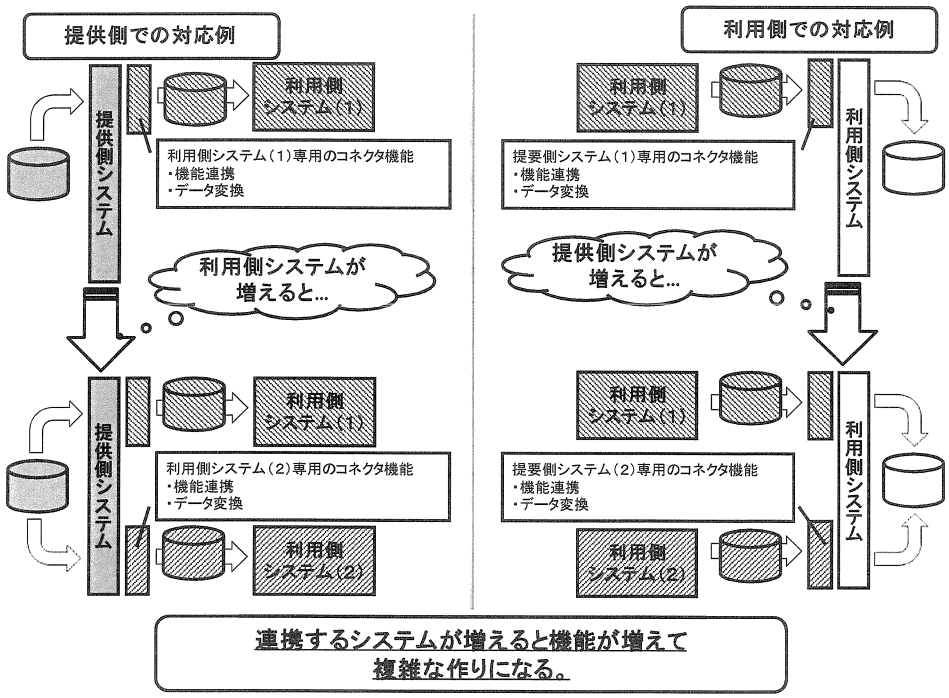


図1 問題点1

現行の情報システムにおいて，あるシステムから他のシステムに対して情報/データを提供するということは,情報/データを複製することと結果的に同じである．場合によっては提供された側で変更や追加が行われ,原本と異なる情報/データが以後流通することもあり，時間の経過とともに複製と移動が増えて責任の所在が曖昧になってしまうことになる．

また,元々の情報/データを加工した結果が一度利用者や連携するシステムに提供されると,ソースになった情報/データに誤りがあっても誤りが誤りとされないまま流通し続けるおそれがある．前述のように複製されたり,加工された結果が流通するので,オリジナル・データにあった誤りや途中での取り扱い方は決して取り除かれることなく残ることになる(図2を参照)．したがって，それらを修正するには多大な時間と手間が必要となる．

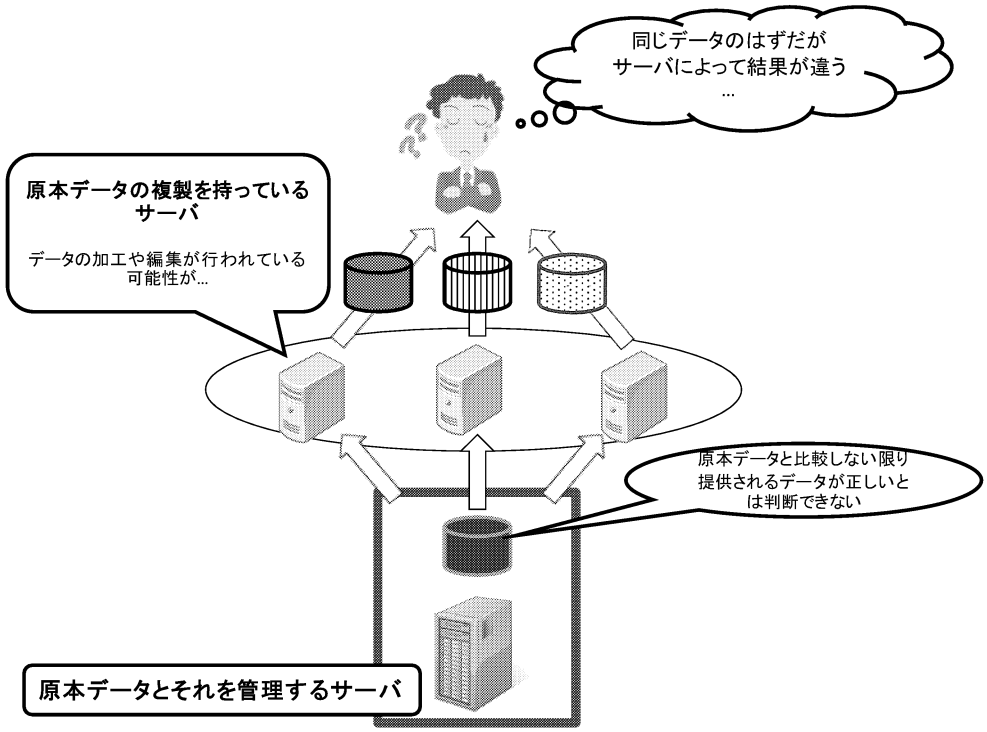


図2 問題点2

3.1.2 解決策となる“リファレンス構造システム”とは

このような従来型情報処理システムにおける問題点は，従来型システムが特定の目的を効率的に達成させるためだけに構築された合目的的なシステムであり，その結果として段階的・階層的なシステム構築が行われることにより,原本の情報/データではなく，コピー，あるいは加工された情報/データを利用することが，何ら疑いなく通常の手段となっていることに起因していると考えられる．

これに対して，リファレンス構造システムは，情報/データの生成・作成からそれら

を加工し結果を利用する一連の処理の流れを生成・作成と加工・利用との間で一旦切り離し，合目的的に一体化されたシステムではなく，状況に応じて柔軟に組み替え再構築を容易にするシステムの在り方を目指している．情報/データを生成・作成して提供する側とその情報/データを加工・利用する側との間のインターフェイスを単純化し，提供側と利用側がより容易に参加できるようにすることで，リファレンス構造システム上で多くのコンピュータシステムが連携しても合理的かつ効率的に情報/データの提供と利用が行える仕組みとなる．

リファレンス構造システム上で，情報/データを提供する側と利用する側とは，固定的な連携相手ではなく，扱う情報/データや目的によって相手が変わることになる．そのため，情報/データについてメタ情報を作成しリポジトリとして蓄積・運用し，必要な情報/データがどこにあるのかを提供する手段や，情報/データを流通させ，それらを受け渡す手順は，リファレンス構造システムとして明確に定義されることになる．

また，提供側と利用側がｎ対ｎの関係になるため，流通する情報/データについてもある程度の規則化が必要である．情報/データの標準化はリファレンス構造システムの目指すところではない．標準化はその策定に多くの時間とエネルギーを必要とし，情報/データとしての柔軟性を犠牲にする．そこでリファレンス構造システムでは“マイクロデータ”と言う考え方を導入し，多くのシステムが受け入れ可能な情報/データの記述を採用することにした．

リファレンス構造システムは，このような従来型の情報処理システムに置き換わるものではなく，従来型の情報処理システムが抱えている欠点を，従来型システムを大きく更新することなしに解決する一つの方法になる．

3.1.3 リファレンス構造システムはどのように機能するか

従来のように情報/データを作成・生成していたシステムでは，その情報/データを扱う特定のシステムを対象とし最適化されていた．リファレンス構造システムにおいては，このようなシステムを「情報所有者（無体物である情報/データには所有権は発生しないが，便宜的に「所有」と表記する）」と定義する．「情報所有者」には，特定のシステムや利用者に情報/データを提供することを求めない代わりに，自身が所有する情報/データのメタ情報を作成し，公開することを求めることになる．情報所有者が作成するメタ情報は「情報案内者」と呼ぶリポジトリ・システムに蓄積され，「情報利用者」に対するメタ情報検索環境を提供することになる．

「情報利用者」とは，従来型システムにおいて情報/データの加工・利用を行っていたシステムを指すが，リファレンス構造システムにおける「情報利用者」は予め決定されたシステムから情報/データの提供を受けるのではなく，加工・利用に必要な情報/データがどの「情報所有者」の管理下にあるのかを「情報案内者」のリポジトリにあるメタ情報を検索し，メタ情報に含まれる情報/データの〈所在情報〉を取得し，該当する「情報所有者」から直接情報/データの提供を受けることになる．これにより，「情報利用者」は常にその時点における最新かつ正しいと判断された「情報所有者」が管理する情報/データを利用することが可能になる．

「情報所有者」，「情報案内者」及び「情報利用者」の関係をまとめると，表１のようになる．

表 1 リファレンス構造システムを構成するサブ・システム

項番	サブ・システム	概　要　説　明
1	情報案内者	流通するデータに関するメタ情報のリポジトリ・システム. メタ情報の登録, 更新及び検索する機能を情報所有者と情報利用者に提供する.
2	情報所有者	リファレンス構造内で流通するデータを維持･管理しているシステム. 維持･管理対象のデータを生成･作成してそのデータについてのメタ情報(URI 等の所在情報を含む)を情報案内者に登録する. メタ情報に従ってアクセスする情報利用者からの要求に応じて該当するデータを提供する. データの維持・管理方法は情報所有者の任意となる.
3	情報利用者	リファレンス構造から必要なデータを取り出すサブ･システム, あるいは利用者インターフェイス. メタ情報の検索を行って所在情報を取り出し, 1 以上の情報所有者から必要なデータの提供を受ける.

表 1 における各サブ・システムは, 図 3　　に示す関係となる.

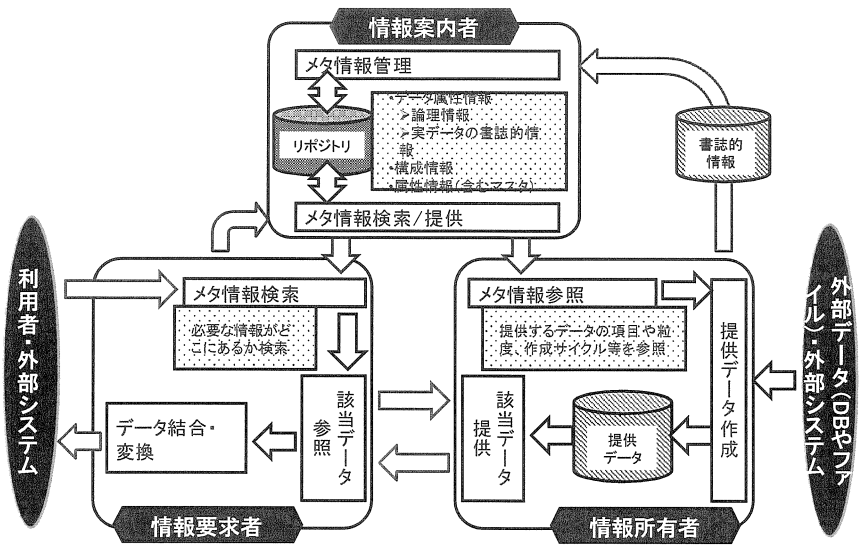


図 3 リファレンス構造システム機能関連図

このように, 情報所有者から提供を受けた情報/データを情報利用者が蓄積することをリファレンス構造システムでは前提としていない. 蓄積するデータは〈所在情報〉のみであり, 情報/データは常に情報所有者から提供を受けることになり, 従来システ

ムのように提供された情報/データを蓄積し複製を増殖させないで必要な時点で一時的に情報所有者から受け取って扱うことになる．従来型の永続的なデータ蓄積とは違うことを示す意味で，情報利用者側には処理の間だけ読み込み可能な情報/データとして存在することから，このような仕組みを「リファレンス構造」と呼ぶことにする．

3.1.4 リファレンス構造システムにおけるマイクロデータの位置付け

リファレンス構造システムを成り立たせるには，情報所有者と情報利用者の間で流通する情報/データがどのようなものであるかを双方であらかじめ確認しておく必要がある．一般にこのような場合には情報/データの標準化を図り，標準に準拠することで複数のシステムを連携させることになり，システム全体に対する情報/データの標準化を進めようとすると状況によっては情報所有者や情報利用者のシステムを根本から作り直すことになる．そこで，リファレンス構造システムでは，情報所有者から送り出された時から情報利用者が受け取るまでの間だけ「マイクロデータ」と呼ぶ方式で情報/データを維持する．一般的な標準化のように情報/データの形式や構成・構造等を厳密に定義するのではなく，含まれているべきデータ項目やデータ・タイプ，データ粒度を申し合わせたものが「マイクロデータ」となる．紙に記述された見積書であれば，それを作成した会社毎に形式や項目の呼び方，レイアウトが異なっていてもそれらを組み替えて理解し，比較することができるが，マイクロデータはそのような自由度を持ったデータ記述の方法である(図４を参照)．

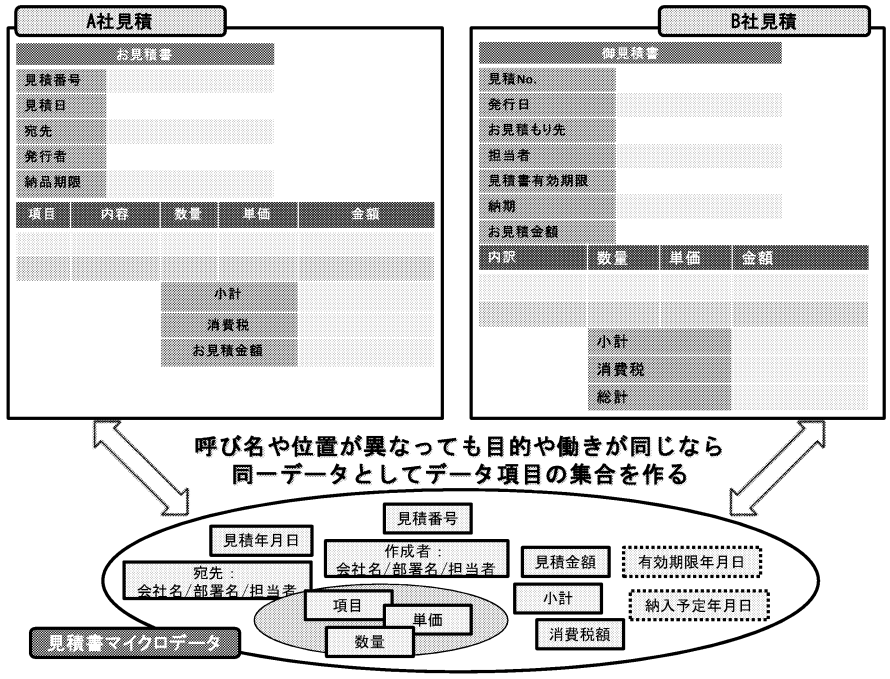


図４ マイクロデータの考え方

すなわち，マイクロデータとは，リファレンス構造システムを用いたデータ参照のために定義されたデータであり，情報利用者が求める情報を構成する最小単位のデータとなる．

実際には，情報利用者の利用目的によりマイクロデータの粒度はその都度定義されることになるため，マイクロデータの実態は情報利用者の要求によって異なることになる．

マイクロデータは，最小単位のデータであることから，必要に応じてデータを取りまとめることで，より上位の単位のデータを生成することができる．そのため，情報所有者が情報利用者の要求ごとにその都度情報を生成することがなくなり，様々な用途でデータを利用することができるようになる．

また，マイクロデータのデータ項目やデータ・タイプ，データ粒度等はメタ情報化されて管理され，情報所有者や情報利用者に提供される．情報所有者はその情報に従い独自の形式や媒体，管理機能に格納されたデータをマイクロデータに変換し，情報利用者はマイクロデータを利用者のシステムで使用できるように変換して処理することになる．

表 2 マイクロデータの構成

名前	説明
データ項目	情報所有者から提供されたデータ内に含まれている項目の名称を示す. 例）都道府県名，年月日　・・・
データタイプ	情報所有者から提供されたデータがどのような形式で保存されているかを示す. 例）ファイル形式：XML 形式 A 項目　　：文字列
データ粒度	情報所有者から提供されたデータがどのような単位で保存されているかを示す. 例）時間軸：年度別，月別，日別 地域軸：全国，都道府県，市町村

3.2 リファレンス構造システムとマイクロデータの特徴とメリット

3.2.1 オリジナル・データへのアクセスが基本

リファレンス構造システムの中で流通するデータは，基本的にそのデータを作成・生成したシステムの管理下に留まる．そのデータを利用したい情報利用者がその都度データ提供を情報所有者に求め，データの提供を受けることになる．これにより，情報利用者は常に必要な時点でのデータを情報所有者に求めるので，その時点での最新結果を得ることが可能となる．

情報所有者が提供するデータにオリジナルであることを明確にする（責任の所在を明確にする）ために電子署名[3]を付与する等の処理を加えれば，データの勝手な一人歩きを回避することができる．

なお，リファレンス構造システムは PKI のみを前提にするものではなく，PKI に替わるより正確で効率的なデータの証明手段が普及すれば，こちらを用いてデータの真正性を保証することになる[4]．

3.2.2 情報所有者による直接的な情報/データ提供

従来型システムのようにバケツリレー方式でデータの保管・蓄積を行った場合には，オリジナルのデータに誤りがあっても派生データが数多く存在する可能性があり，全のデータを修正することは非常に困難となり，全てが修正できたかを確認することは不可能といえる．また，従来型システムでデータを受け取る側にとっては，データのトレースができないために，オリジナルのデータにアクセスできているのかが不明であり，修正・編集・改竄(数字であれば四捨五入や切り捨て，単位変更等)が加えられているのか否かを判断することはできない(図5を参照)．

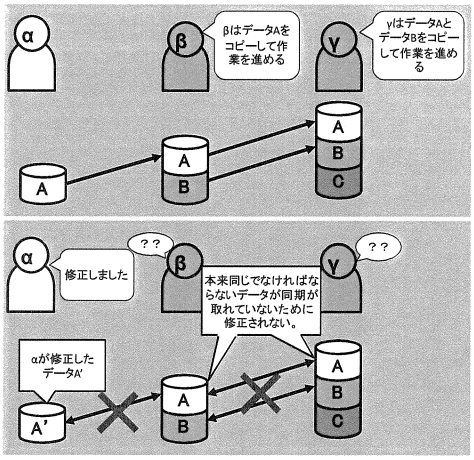


図5 従来型システム(バケツリレー)におけるデータの流れ

データの発生元がオリジナルのデータを直接提供する場合には，データにまつわる上記のような懸念はなくなる(図6を参照)．

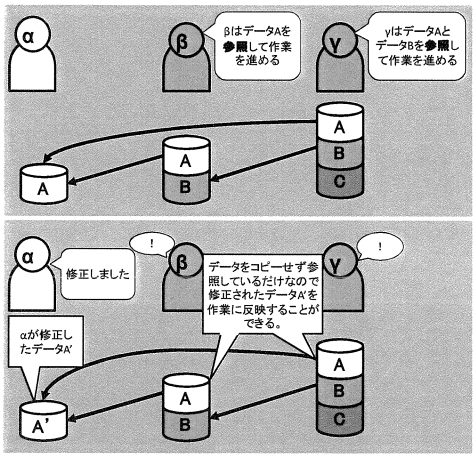


図6 リファレンス構造システムにおけるデータの流れ

3.2.3 情報所有者によるアクセス制限

ネットワーク上でオリジナル・データからの派生データが数多く流通すると，情報所有者は自身の発信したデータがどこでどのように扱われているのかを知ることができず，情報所有者にとって不本意な扱いを受ける可能性がある．情報所有者が情報利用者に直接データを提供するのであれば，誰に何を提供したのかが明確であり，項目単位や項目の組み合わせに対して情報利用者の権限等に応じたきめ細やかなアクセス制限を加えることが可能となる．

3.2.4 マイクロデータによる柔軟なデータ整合性

マイクロデータは構成項目やデータ形式等もメタ情報として管理し，従来のような標準化を避けることで柔軟なデータ整合性を目指すものである．提供されるデータから構成項目の位置やサイズ，並び順等の属性を一旦分離し，利用者側に到達した時点で分離された属性を利用者側の規則で再構成する．したがって，マイクロデータを使用する上では最低限の決め事だけが必要で，厳密な規格化・標準化を行う必要はない．

提供する情報/データの蓄積機能や蓄積する情報/データのスキーマを変更することなく，その独自形式の情報/データをマイクロデータに変換する機能（フィルタ機能）を追加することでマイクロデータとリファレンス構造に対応できることになる．

情報利用者側においても同様であり，現在リレーショナル DB 等から情報/データを取得する機能を使用することでリファレンス構造システムからマイクロデータとして情報/データを取得し，マイクロデータを独自形式の情報/データに変換する機能（フィルタ機能）置き換えることで対応することができる．

マイクロデータは緩やかさを持つデータ定義方法で，その緩やかさを利用することで既存システム等に対する大幅な変更等を加えることなく，リファレンス構造システムの一部として拡張を行うことができる．

4 次世代情報処理基盤を支える技術

4.1 データ記述・管理

マイクロデータは，マイクロデータとリファレンス構造システムに関する規約に基づいたメタ情報の記述により管理される．メタ情報は，表 3 のようにデータの概要を示す「論理的属性情報」，データのファイルレイアウト情報などを表す「データ構成情報」及びデータを特定するための情報を表す「データ属性情報」で構成される．

表 3 マイクロデータの構造

	概要	種別	作成者	作成タイミング	保管者
①論理的属性情報	データを何のために，どのように作成するかを記すもの	メタ情報	情報所有者	データ要求時	情報案内者
②データ構成情報	データをどのような項目で作成するかを記すもの	メタ情報	情報所有者	データ要求時	情報案内者

③データ属性情報	データの所在，意味を表すもの	メタ情報	情報所有者	データ作成時	情報所有者
④実体データ	データそのもの	実体データ	情報所有者	データ作成時	情報所有者

4.1.1 論理的属性情報

論理的属性情報とは実際のデータに対して，どのような目的で作成されていて，どのような期間，ルールで作成するかを記したものである．

この論理的属性情報は，データの管理者，つまりデータの発生を促す必要のある人間が定義することになり，情報所有者に対する要求仕様ということもできる．

論理的属性情報は，その意味をコンピュータにも理解できることを目指すため，メタ情報記述に使う語彙として，既に多くの分野で利用されている DublinCore[5]の考えを採用することになる．

【論理的属性情報に含まれる内容】

論理的属性 ID
名称
マイクロデータ名称
別名
作成者
主題
説明
提供者
関連組織等
日付
作成日
利用可能日
正式発行日
更新日
登録期限
データタイプ
リソース識別子
関連情報識別子
地理的な範囲及び対象
時間的な範囲及び対象
権利

4.1.2 データ構成情報

マイクロデータをどのようなファイルフォーマット，項目で構成するかといった情報を記載する．情報所有者に対するデータ要求仕様ともいうことができる．

【データ構成情報に含まれる内容】

ファイル定義
├名前
├形式
├関連情報識別子
├論理的属性識別子
├データ構成識別子
├レコード
├項目識別子
└項目表示順序
　├データ項目
　├識別子
　├項目定義
　├データ種別
　├最大バイト数
　├キー項目
　├省略可能
　├カラム数
　├項目の単位
　└カテゴリ識別子

カテゴリ情報
├カテゴリ情報
├カテゴリの説明
├カテゴリ名

├カテゴリラベル
└カテゴリコード

4.1.3 データ属性情報

実体となるマイクロデータを特定するためのメタ情報であり，論理的な属性情報とは異なり，データを公開するタイミングで情報所有者によって作成されるものである．実体のマイクロデータとは必ず対の関係で存在する．

データ属性情報は，その意味をコンピュータにも理解できることを目指すため，メタ情報記述に使う語彙として，既に多くの分野で利用されているDublinCoreの考えを採用することになる．

【データ属性情報に含まれる内容】
- マイクロデータID
- 名称
- マイクロデータ名称
- 別名
- 作成者
- 主題
- 説明
- 提供者
- 関連組織等
- 日付
- 作成日
- 利用可能日
- 正式発行日
- 更新日
- 登録期限
- データタイプ
- リソース識別子
- 関連情報識別子
- データ構成識別子
- 地理的な範囲及び対象
- 時間的な範囲及び対象
- 権利
- 更新履歴情報
- データ検証用ハッシュ値

4.2 利用するプロトコル

4.2.1 サービス連携

リファレンス構造システムは広く一般に使用される情報/データの流通基盤となることを目指しているので，情報利用者にサービスを提供する情報案内者と情報所有者の機能は疎結合な環境を想定し，Webサービスとして実装することになる．

Web サービスと言えば“SOAP”による構築が想定できるが，WSDL でサービスを記述し厳密なインターフェイスが実現できる分だけ全体に統一された環境が必要となる．リファレンス構造システムでは，Webサービスでも RESTful 型[6]でのサービス連携を想定している．複数の情報利用者と複数の情報所有者，そしてそれらと情報案内者とが相互に連携するシステムとなるので，インターフェイスの厳密さを求めるよりも仕組みの理解が容易で実装時の相互のテスト等が行いやすいRESTでのWebサービスが望ましいと考えている．

RESTful 型の Web サービスでは，サービスが提供するリソースを URL として記述する．その URL に対して GET，PUT，DELETE の HTTP で定められたメソッドでアクセスを記述する方法は単純化されていながらも，引数等と組み合わせることで十分に高度な利用目的を達成することが可能となる．

情報利用者側から情報所有者を見ると，情報利用者が求める情報/データを持っている情報所有者は複数存在することが考えられる．リファレンス構造システムではどの情報所有者に情報/データを要求するかは，情報案内者から取得した所在情報に従って同じ手順，プロトコルでリクエストするので，一つの情報所有者にリクエスト

しているのと見かけ上は同じになる．逆に情報所有者から見た場合でも，アクセス制限等の問題を別にすれば，見かけ上の情報利用者は一つになる．リファレンス構造システムはこのような単純化された，システム全体規模が自在に変化する疎結合なサービス連携を実現する．

4.2.2 リポジトリのクラスタリング

情報案内者は，リファレンス構造システム全体が利用するメタ情報を集積し，その情報を全ての情報利用者と情報所有者に提供する．リポジトリとして集積するメタ情報は，リファレンス構造システム内へ提供するデータが生成される都度，情報所有者から登録され，メタ情報が追加されることになる．

リファレンス構造システムの中で情報案内者は，その役割からすると以下のようなことが求められる．

・高いスケーラビリティと柔軟性
・負荷分散と信頼性
・高いアベイラビリティと短い応答時間

これらを実現するために，情報案内者は複数のノードで構成されるクラスタとして動作することを想定している．その上で全体を管理するマスターノードを持たないようにすることで，単一故障点をなくすことができる．したがって，ノードの追加削除にはシステム停止やデータの再配置が不要になることから，運用状況に応じてシステム規模を柔軟に調整することができる．

一般的にクラスタ・システムでは，クラスタを構成するノードリストをノード間で共有しなければならない．共有方法としてゴシップ・プロトコルと呼ばれるプロトコルを利用して，各ノードの状態等を共有し，ノードの故障や追加等に対応する．

また，ノード間でデータを分担して管理するので，負荷分散にも対応できると考えている．

4.3 メタ情報

リファレンス構造システムの中で情報所有者から情報利用者に提供される情報/データには必ずメタ情報があって，それらは情報案内者のリポジトリ内で管理されることになる．

流通する情報/データには“論理 ID”が付けられる．論理 ID は見積書，請求書，住所変更届等の概念的，論理的な情報を示し，それらは“論理的属性情報”と“データ構成情報”のメタ情報として構成される．

“論理的属性情報”は，インターネット上のリソースに関する情報を記述して有用な情報の探索・発見に役立てる目的で作られた Dublin Core 基本語彙に従って記述する． Dublin Core 基本語彙は，専門家でなくとも簡単に記述できることを目指して，簡易なメタ情報を作成するという意図から作られたため，必ず記述しなければならない必須項目や各項目の記述順序は無く，同一項目を複数回使用することも自由であり，柔軟で緩やかな方針がリファレンス構造に適している．

“データ構成情報”はそれが示す情報/データがマイクロデータとして実体化された際に，構成する項目の集合について，下位の集合や，その項目の名前やデータ・タイプ(日付，数値，文字列等)等について定義を行う．

“論理的属性情報”と“データ構成情報”はそれが示す情報/データの実体が存在しなくてもリポジトリ内で管理される．一つのリファレンス構造システム内において，バージョンは別とすれば，ある論理 ID に関連づけられる“論理的属性情報”と“データ構成情報”は一セットとなる．

これに対して，論理 ID“見積書”に対応したデータ実体が発生すると「2009 年 4 月 1 日に A 社が B 社宛に発行した台湾産バナナの見積書」のように具体的な属性情報がメタ情報として発生し，情報所有者から情報案内者に登録される．これを“データ属性情報”と呼ぶ．“データ属性情報”はデータ実体の書誌的な情報とそのデータ実体にアクセスするための“所在情報”で構成され，“論理的属性情報”と同様に Dublin Core 基本語彙で記述される．

これらのメタ情報は情報案内者のリポジトリの中で，RDF 形式で管理し，RDF クエリ言語の一種である SPARQL（SPARQL Protocol and RDF Query Language の略）で検索/操作を行う[7]．現時点ではメタ情報を RDF 形式で記述し SPARQL で操作することを想定しているが，技術動向では他の技術要素で対応する可能性もある．

4.4 サービス提供/利用手順

リファレンス構造システムでは大きく以下の 2 段階でサービス提供/利用を行ことになる．

第 1 段階：情報/データ提供の準備

①情報所有者側に存在するシステム内で情報/データが生成・作成される．

②情報所有者はその情報/データをリファレンス構造システム上で提供するのであれば，それらについてメタ情報（データ属性情報）を編集する．

③メタ情報は情報所有者によって情報案内者に登録される．

④情報案内者は登録されたメタ情報を論理 ID に関連付けてリポジトリ内に管理する．

論理 ID で概念的・論理的な情報/データを識別するために，“論理的属性情報”と“データ構成情報”をメタ情報として 1 セットずつ持つが，同じ論理 ID に関連づける“データ属性情報”は実体データ毎に作成されるので，複数存在する．また，実体データに誤りが発見されて実体データ上を修正した場合には，情報案内者が管理するデータ属性情報にそれを作成した情報所有者によって修正の履歴が記録される．これにより，情報利用者はメタ情報を見るだけで情報/データの変遷を知ることができることになる．

第 2 段階：情報/データ提供の実施

①情報/データを必要とするユーザやシステムは，情報利用者機能を利用して必要な情報/データの検索条件を情報案内者に送り，該当する情報/データを検索する．

②情報案内者は受け取った条件でリポジトリ内を検索し，検索結果を情報利用者に応答する．

③検索結果が不満足な場合，情報利用者側では検索条件を追加・変更し①からの手順を繰り返す．

④情報利用者は情報案内者から得た所在情報のリストに従い，それぞれの情報所有者へ情報/データ取得として以下の作業を行う．

1）情報利用者は所在情報を元に情報所有者と接続し，情報/データを情報所有者に要求する．

2）必要があれば情報所有者は情報利用者の権限，要求された情報/データのアクセス制限に従って要求の受け入れを判断する．要求を受け入れられない場合はその旨を応答して切断する．

3）情報所有者は要求されている情報/データをローカル・システムから取り出し，その情報/データをフィル

タ処理でマイクロデータに変換して情報利用者に提供する．また，現在要求している情報利用者に提供できない項目が含まれている場合，その項目をマイクロデータに含めないことも可能となる．

4) 情報利用者はマイクロデータを受け取った後に情報所有者とのコネクションを切断する．

5) 受け取ったマイクロデータはフィルタ処理を経ることで，情報利用者側で取り扱うことができる情報/データに変換される．

⑤全所在情報を処理したら取得した情報/データを情報利用者機能の利用者であるユーザやシステムに引き渡す．

フィルタ処理は情報所有者と情報利用者の双方で機能するが，どちらも情報所有者や情報利用者が独自に扱う情報/データとマイクロデータ間での変換を主な機能としている．独自の情報/データとマイクロデータ，あるいはマイクロデータと独自の情報/データとの間での変換は原則として可逆的な変換であるが，変換前より変換後の方のデータ粒度が粗くなる（例：日次が週次，課単位が部単位）ような集計等の加工が伴う不可逆的な変換も必要に応じて実施する．

4.5 セキュリティ

リファレンス構造システムはネットワークでつながった利用者，システム，サービス間で情報の提供と参照を行うためのものであり，情報/データを安心して安全に利用するためにはセキュリティ機能が不可欠となる．ここで求められるセキュリティ機能とは，情報/データを提供するシステムや利用する側の保護を目指すだけのものではなく，ネットワーク上で流通する情報/データの様々な意味での正しさを保証することが重要となる．

セキュリティ機能には，コンピュータやネットワークを利用するシステムとしてその構成や働き方として実現されるものと，これらのシステムによって利用可能となる業務やアプリケーションとして実現されるものに大別することができる．

こらマイクロデータとリファレンス構造システムそのものが，ネットワーク上に分散/散在している複数のシステム等を緩やかに結合して情報/データ共有を目指すものであり，完全に閉じた環境での利用でない以上，連結・連動を容易にするためにも，セキュリティ機能に限らず，公開され広く認識された技術を採用する．開かれた技術を採用することは，システムがヒトから見えないところで何を行っているのかをそのシステムの周囲に公開することを意味するものであり，これもセキュリティ機能の一部として考えることができる．

以下，リファレンス構造システムとマイクロデータを利用したシステムとして実現されるべきセキュリティ機能について示す．

4.5.1 暗号化通信

リファレンス構造を構成するシステム（情報利用者，情報所有者，情報案内者の各役割）の間で通信を行う場合に通信を暗号化して盗聴や改ざんを防ぐことを目的とし，SSL を利用する．

4.5.2 真正性の担保

電子データに PKI を利用した電子署名を付与してその真正性を担保することが一般的になってきていることから，リファレンス構造システム上で提供・参照の対象となるマイクロデータに対しても電子署名を付け，その情報/データと提供者の関係を明確にするだけでなく，提供者が保証する範囲を示すことでデータの真正性を担保し，そ

れらを利用したシステム信頼性を高める．

リファレンス構造を利用するシステムの中では，最終的な目的である情報所有者から情報利用者に対する情報/データの提供に対してのみ真正性を求めるのではく，情報所有者から情報利用者に至る過程で流通するメタ情報に対しても同様とする．隣り合うシステムや機能間での情報/データの交換/参照においても電子署名を用いた真正性の確認を行い，それらを積み上げることで全体としてのシステム全体が真正性の担保を有するようになる．

4.5.3 データ確定の保証

情報/データが情報所有者側の事情により修正・変更が行われ値や記述が変更されることが考えられるが，このような場合には，修正前後のデータにそれぞれメタ情報を持たせ，実体としては異なるデータとして管理することになる．つまり，その情報/データにアクセスした情報利用者は，メタ情報に含まれる更新履歴情報や更新日を利用することで，修正前あるいは修正後の任意の情報/データのデータにアクセスすることが可能になる．

リファレンス構造システム上で提供することを前提に公開された情報/データは，公開以降は，情報利用者がいつでも同じ情報/データにアクセスできることを保証する．物理的なデータの書き換え禁止は情報所有者側のシステムにおける実装に依存するが，データが変更されていないことを情報利用者に保証する方法として，公開対象の情報/データのハッシュ値を情報案内者が管理するメタ情報に含める．これによって情報利用者は受け取った情報が必要なものであるのかを検証することが可能となる．

4.5.4 アクセス制限

アクセス制限として想定される機能には次の 2 種類がある．

(1) 情報/データの秘匿・非提供
公開される情報/データに個人情報等の秘匿すべき内容を含んでいた場合に，提供する情報/データの該当部分にマスクをするか，該当する情報/データを提供内容に含めない等の対処を行うことになる．

(2) 機能に対する制限
提供する情報/データを情報利用者の権限や立場に併せて制限し，参照可能な範囲を制御する情報所有者/提供者の機能によるアクセス制御を行う．

リファレンス構造システムは組織や機構を跨る大規模な利用環境を前提としている．したがって，規則や考え方の異なる組織や機構を統合する認証機構の構築が非常に困難なものであることから，一元的な利用者認証に基づくものではない．

また，リファレンス構造システムの機能から情報/データに至る範囲の変動に対応するには，個人を特定するより個人に割り当てられる権限や属性(立場)により管理する方が現実的である．

5 まとめ

5.1 今後必要となる取り組み

データの粒度と意味を定義するだけの緩い標準化といえ，情報所有者にはメタ情報の付与をはじめとするこれまでにない作業が生じるため，現行のカスタマイズにより特定の業務を実施するためだけに特化したシステムと比較すれば，業務を実施するという限られた範囲では「余計」な作業が追加されることになる．しかし，データが共有化されていないことによるデータ非同期をはじめとする様々な社会的リスクを考慮

すれば，社会全体にとってメリットとなる方向にデータのあり方を切り替えることが必要であることから，社会的ニーズと個々の現場における業務担当者との合意が不可欠な要素となる．こうした合意を積み重ねて，緩い標準化の内容を具体化していく努力が情報所有者と情報利用者双方に求められることになる．

5.2 システム構築時の検討

リポジトリ・システムはネットワーク上に複数存在し，あたかも一体であるが如く機能することが前提となる．そのためには，各リポジトリ・システム間でメタ情報等に関する同期が取れていなければならないが，リポジトリ・システムは分野(業務)単位で構築されるであろうことを想定すると，分野(業務)横断的な取り決めによりデータ更新のタイミングを設定する必要がある．

マイクロデータに関しても，そのオリジナル・データは情報所有者のコンピュータに記録されていることから，情報利用者の要求に応じたマイクロデータを情報所有者のコンピュータから直接生成することはセキュリティ上問題がある．したがって，マイクロデータを生成させるための公開用コンピュータが必要となるが，この公開用コンピュータのデータと情報所有者のコンピュータにおけるオリジナル・データとの同期を確保する必要がある．業務によっては同期のタイミングにばらつきがあるとデータして使用することができないこともあるため，同期のタイミングについては，情報利用者と情報所有者が十分議論する必要がある．

5.3 メタ情報の管理

どの粒度のオリジナル・データにどのようなメタ情報を付与するのが最も効率的であるかを決めるには，情報所有者と情報利用者とによる議論の積み重ねによるか，あるいは実際の利用環境での使用状況に基づくデファクト化によるか，複数の選択があり得るが，いずれにせよ，得てして融通のきかない“上からの”標準化ではなく，多くの参加者による実際の使用を踏まえ，自然に収斂するような流れによる“標準化”が現実的かと考える．ただし，“標準化”された事項は，規約等の文書にて関係者に共有されるべきである．

議論の結果確定したメタ情報にせよ，実際の利用環境でデファクト化されたメタ情報にせよ，メタ情報を記録・維持管理し，社会全体で共有するための機構を創設することが必要である．この管理機構は，リポジトリ・システムとは異なり，重複するメタ情報や「業界標準」という名の方言の発生を防ぐために唯一の機構であることが必要となる．

5.4 マイクロデータの定義と公開

オリジナル・データそのものの粒度や意味が情報利用者と情報所有者との合意形成の過程で紆余曲折したとしても，マイクロデータはリファレンス構造システムにより電子データを共有するための基盤であることから，この合意形成プロセスとは非同期でマイクロデータを定義することができる．

マイクロデータに関する定義を公開し，その後研鑽していくことで，リファレンス構造システムによる電子データの共有化を推進していくことが重要である．

5.5 セキュリティ

個人情報に関しては，直接的な個人情報はアクセス制御で権限なきユーザに対して参照不可とすることで対処できるが，直接的な個人情報でない間接的なデータを組み合わせて個人を特定することも可能となることから，間接的データのどこまでを個人

情報に準じて扱うかを整理する必要がある．

個人が特定されるという面だけを強調することで，間接的データも含め一切の個人情報に関するデータの参照を不可とした場合には，データの有効活用に逆行する場合もあることから，個人情報に関する扱いは慎重かつ大胆に検討を進める必要がある．

また，リファレンス構造システムにおいては，アクセス制限に権限認証や属性認証を行うことを視野に入れているが，現時点において十分に普及したそれらのサービスやサービス基盤は見あたらない．今後，リファレンス構造システムだけではなく様々な場面で権限認証や属性認証は求められることは想定されるので，権限認証や属性認証の普及やサービスの提供が待たれる．

5.6 メタ情報の“進化”に対する対応

複数のメタ情報が組み合わさり，元のメタ情報とは異なる意味のメタ情報が新たに生成されることも想定することができる．こうした場合，組み合わさった後のメタ情報を新規に扱うためのルール策定が必要となる．

参考文献

［1］アトムデータによる国民経済計算の推計　高島史郎　情報知識学会誌 18(3), 260–279, 2008–10–20

［2］データ連携を容易にするリファレンス構造とマイクロデータ技術
http://www.kantei.go.jp/jp/singi/it2/nextg/meeting/dai6/siryou5.pdf

［3］電子署名の仕組み　財団法人　日本情報処理開発協会　電子署名・認証センター
http://www.ecom.or.jp/esac/intro/shikumi.html

［4］PKI 関連技術解説
http://www.ipa.go.jp/security/pki/

［5］The Dublin Core（R） Metadata Initiative
http://dublincore.org/

［6］Leonard Richardson，Sam Ruby（監訳：山本陽平，訳：株式会社クィープ），RESTful Web サービス，オライリー・ジャパン，2007

［7］SPARQL Query Language for RDF
http://www.w3.org/TR/rdf-sparql-query/

2010年度論文賞記念講演

大学における非文献コンテンツ公開のための共通プラットフォームの開発 --- 非文献コンテンツのための可視性と保守性に優れた学術情報リポジトリの構築 ---

Development of Common Platforms for non-Bibliographic Contents Disclosure in University --- Construction of an Academic Resource Repository Excellent in Visibility and Maintainability for non-Bibliographic Contents ---

高田良宏[1*], 笠原禎也[1], 西澤滋人[2], 森雅秀[3], 内島秀樹[4]
Yoshihiro TAKATA, Yoshiya KASAHARA, Shigeto NISHIZAWA,
Masahide MORI and Hideki UCHIJIMA

1 金沢大学総合メディア基盤センター
Information Media Center, Kanazawa University
E-mail: yosihiro@kenroku.kanazawa-u.ac.jp, kasahara@is.t.kanazawa-u.ac.jp
2 金沢大学自然科学研究科(現在, ソラン 株式会社)
Graduate School of Natural Science and Technology, Kanazawa University
(Currently, SORUN CORPORATION)
3 金沢大学人間社会学域人文学類
School of Humanities, College of Human and Social Sciences, Kanazawa University
4 金沢大学情報部情報企画課
Information Infrastructure Service Division, Kanazawa University
*連絡先著者 Corresponding Author

我々は，大学内に蓄積している学術情報の内，その公開が進んでいない非文献コンテンツに焦点をあて，それらを学内外に公開するための指針となるべく，既存情報インフラ上で運用可能な公開手法を考案し，共通プラットフォームとして提供することを目指している．その一環として，本研究では，非文献コンテンツをリポジトリ化する場合の問題点を整理し，その解決のため，メタデータの互換性，異種コンテンツの共存，大規模コンテンツの管理，コンテンツに関する位置情報を用いた可視化などを考案した．そして，既存リポジトリプラットフォームのDSpaceを拡張し，可視性と保守性に優れた汎用性の高い学術情報リポジトリの構築を行った．本講演では，研究の背景と構築したシステムの概要を述べる．

1 はじめに

非文献コンテンツのための可視性と保守性に優れた学術情報リポジトリの構築[1]は，我々が，大学における非文献コンテンツ公開のための共通プラットフォームの開発の一環として行ってきた研究で，その成果をまとめ情報知識学会誌に投稿させて頂いたものである．この度，当論文が情報知識学会論文賞に選ばれたことは，誠に光栄なことであり，ご指導ご協力頂いた関係者および推薦委員の先生方に深く感謝申し上げる次第である．

情報化が進む今日，大学は貴重な学術情報を蓄積するだけではなく，世界に向けて発信することを求められている．大学が公開している学術情報は，学内に蓄積している膨大な学術情報の一部であり，多くの学術情報は公開に至らず，学内に死蔵しているケースも少なくない．それに対して，昨今，各大学では学術論文，紀要，研究報告などの学術情報を登録し，機関リポジトリとして公開している．一方，写真，動画，音声などのコレクションや実験観測データをはじめとした文献以外の学術情報は，教育・研究上有用なものでありながら対象外とされている場合が多い．

このような背景のもと，本研究では，学内に蓄積している学術情報の内，公開が進んでいない写真，動画，音声などのコレクションや実験観測データに焦点をあて，それらを学内外に公開するための指針となるべく，既存情報技術を駆使した，既存情報インフラ上で運用可能な公開手法を考案し，共通プラットフォームとして提供することを目指した．

本講演では，まず，研究全体の目的にあたる，大学における非文献コンテンツ公開のための共通プラットフォームの開発について述べる．続いて，今回受賞対象となった，非文献コンテンツのための可視性と保守性に優れた学術情報リポジトリの構築について，その概要を述べる．

2 非文献コンテンツ公開のための共通プラットフォームの開発

本章では，非文献コンテンツのための可視性と保守性に優れた学術情報リポジトリの構築を行うに至った背景であり，研究全体の目標にあたる非文献コンテンツ公開のための共通プラットフォームの開発について概説する．

2.1 非文献コンテンツ

大学内に蓄積している学術情報の内，学術論文や紀要，研究報告書などの文献系のコンテンツを文献または文献コンテンツと呼ぶのに対し，文献系コンテンツ以外の写真，動画，音声，教材などのコレクションや実験観測データを非文献コンテンツと呼ぶ．

2.2 学術情報公開モデル

本研究では，大学内に蓄積している学術情報の内，その公開が進んでいない非文献コンテンツに焦点をあて，それらを学内外に公開するための指針となるべく，既存情報インフラ上で運用可能な公開手法を考案し，共通プラットフォームとして提供することを目指した．共通プラットフォームとは，公開手法を汎用的に実装したシステムで，そのままの形で，あるいは，簡単な設定や簡単な追加実装で再利用できる，いわゆる「使い回し」が可能なシステムである．

まず，非文献コンテンツを学内外に公開する際の問題点を洗い出し，それらを解決するために，大学内のすべての種類の非文献コン

テンツを網羅する学術情報公開モデルを作成した．このモデルでは，既存情報技術で構築可能であり，既存情報インフラ上で実運用可能な「学術情報リポジトリ」，「Web-DB管理システム」，「データ配信システム」の3つの公開手法で，大学内のすべての非文献コンテンツの公開をカバーするものとした．図1は，作成した学術情報公開モデルを大学の現状に即した形とし，共通プラットフォームを配置したものである．

2.3 各共通プラットフォームの概要

作成したモデルに基づき，それぞれの公開手法の問題点を改善し，「非文献コンテンツに対応した学術情報リポジトリ」，「多様なアクセス制限に対応したWeb-DB管理システム」，および，「汎用データフォーマットを用いたデータ配信システム」として設計を行った．実証運用では，当大学の研究室で実際に蓄積している複数種の非文献コンテンツを用いて，公開システムとして構築・運用し，共通プラットフォームとしての有効性と汎用性を確認した．以下に，開発した公開システムの概要を述べる．

【非文献コンテンツに対応した学術情報リポジトリ】：非文献コンテンツの中でも，写真，動画，音声などのコレクションを公開するための共通プラットフォームとなるべく，既存プラットフォームの問題点を改善し，可視性と保守性に優れた汎用性の高い学術情報リポジトリの構築を行った[1]．この部分に関しては，3章で詳細を述べる．

【多様なアクセス制限に対応したWeb-DB管理システム】：公開用Web-DBシステムの一元的な管理・公開を可能とする共通プラットフォームの提供を目的とし，ユーザおよび公開用Web-DBシステム配下のデータベースごとに，きめ細かなアクセス制限を行うことができるWeb-DB管理システムの開発を行なった[2,3]．実証として，地球環境観測データに適用し，地球環境データベースシステムを構築した．詳細は参考文献2,3を参照されたい．

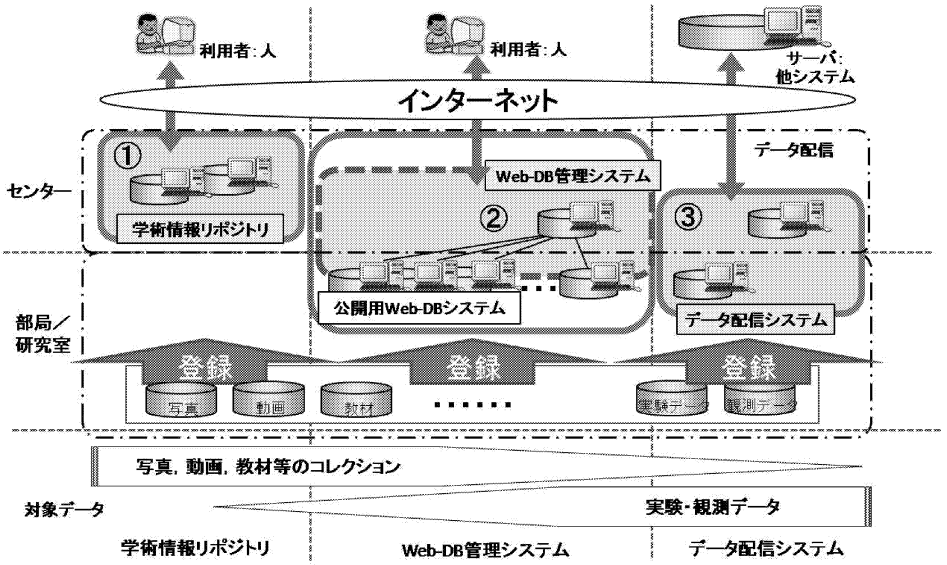


図１　学術情報公開モデル

【汎用データフォーマットを用いたデータ配信システム】:自然科学系の実験観測データを容易に利活用できる形で公開することを目的として,相互配信環境モデルを提案,さらに,その実現のために地球環境観測データに適用し,自然科学データアーカイブスシステムの基本構成の開発を行った[4]. 詳細は参考文献4を参照されたい.

3 可視性と保守性に優れた学術情報リポジトリの構築

3.1 設計方針

課題1:先行する文献リポジトリに関しては,Dublin Core[5]（以下DCとする）の考え方に沿った記述法が確立されつつあるが,非文献コンテンツに対しては,それらがもつ多様で専門的な情報をどのようにDCで定義するかが不明確である.

課題2:保守・管理面から考えた場合,管理者が保有する複数の性質の異なったコンテンツ（異種コンテンツ）をどのように管理するか,数千件,数万件以上におよぶコンテンツをどのように分類,登録・保守するか,などの問題がある.さらに,コンテンツ管理者が必ずしも情報技術の専門家であると限らないという問題もある.このため,情報技術に関して専門外である管理者でも,リポジトリの保守管理ができる仕組みの導入を検討する.

課題3:非文献コンテンツは,一般に文字情報を持っていないので,文献コンテンツに比べ利用者に対する検索性が低いという問題がある.一方,非文献コンテンツには,発掘地,所蔵地をはじめとした地名などの情報を持つものが多い.そこで,利用者に対する検索性を高めるため,地名などの地理的位置情報を得て,より視覚的に検索できる仕組みの導入を検討する.

課題4:各リポジトリに存在する異種コンテンツ同士を横断的に検索,利用することを想定し,リポジトリ間の連携を行い,これらの情報をどのように相互参照するかという検討が必要である.

以上,課題1～課題4を解決し,非文献コンテンツに対応した汎用性の高い学術情報リポジトリを構築するため,表1に示す開発条件を設定した.また,表1には,課題には挙げていないが,⑤プラットフォームの保守性が高いこと,という項目を加え,既存リポジトリプラットフォームをベースにするという条件を設定した.文献コンテンツでの利用実績があり,保守体制が確立しつつある既存

表1　開発条件表

汎用性の確保
①タデータの互換性が確保できること
⇒ 当該リポジトリ上での詳細な定義と,他リポジトリとの互換性を両立（**課題1**）
②他リポジトリとの連携を行えること
⇒ OAI-PMH[6]プロトコルを活用した他リポジトリとの連携を実現（**課題4**）
保守性の確保
③複数の異種コンテンツの管理を容易に行えること
⇒ 同一リポジトリ（同一システム）に異種コンテンツを容易に共存させる環境を導入（**課題2**）
④多様かつ膨大な数のコンテンツの管理を容易に行えること
⇒ 分類の登録・管理機能,一括登録機能を導入（**課題2**）
⑤プラットフォームの保守性が高いこと
⇒ 既存リポジトリプラットフォームをベースとする
可視性の向上
⑥コンテンツが持つ発掘地,所蔵地などの地理的位置に関する情報を可視化し,視覚的な検索機能を提供すること
⇒ 地理的位置情報とGoogle Earthを連携させた検索機能を導入（**課題3**）

リポジトリプラットフォームを利用することが，保守・管理コストの面で有利と判断したためである．そして，リポジトリプラットフォームのDSpace[7]をベースに，機能を改良，追加する形で開発を進めることとした．

3.2 使用データ

今回は，共著者の森がアジア図像集成として蓄積しているインドの仏像・壁画・遺跡に関する画像を対象にして開発および検討を行った．図2(a)に，蓄積されているコンテンツの例を示す．アジア図像集成は，インドの各地で撮影された画像で，総数は2万件以上に及ぶ．従来，画像のメタデータは，データ管理者の森が，地域，タイトル，所蔵，特徴，サイズ，材質，制作年代などをExcelの表形式で管理していた．また，本システム開発後の検証の目的で，あけぼの衛星の観測データ（図2(b)），e-Learning素材（図2(c)），および，第四高等学校物理機器図録（図2(d)）に適用することとした．

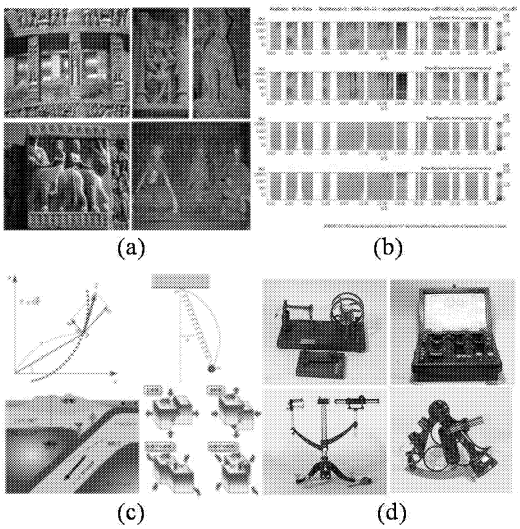


図 2　使用したデータ

【(a)：アジア図像集成，(b)：あけぼの衛星の観測データ，(c)：e-Learning 素材，(d)：第四高等学校物理機器図録】

3.3 システム概要

前述の通り，本システムは，リポジトリプラットフォームのDSpaceをベースにして，機能を改良，追加する形で開発を進めた．開発にあたっては，DSpaceの既存クラスを極力書き換えず，クラスをオーバーライドする形で行い，システムの移植やDSpaceのバージョンアップといった場合にも極力影響が出ないように配慮した．図3は，実装したアジア図像集成の一画面（情報表示画面）である[8]．

以下では，メタデータの互換性の確保，保守性の確保，可視性の向上および他リポジトリとの連携について述べる．

3.3.1 メタデータの互換性の確保

DCの標準メタデータ語彙では，非文献コンテンツの情報を的確に表現することは困難なため，各々のコンテンツに合わせ，メタデータ項目を追加した拡張メタデータ語彙を定義する必要がある．今回，各非文献コンテンツ特有のメタデータ項目の定義にDumb-Down原則[9]を導入した．通常，Dumb-Down原則は，組織間の運用ポリシーの違いなどによるメタデータ項目の差異を吸収するために用いられるが，ここでは，各非文献コンテンツの特性の違いによるメタデータ項目の差異を吸収するために用いた．

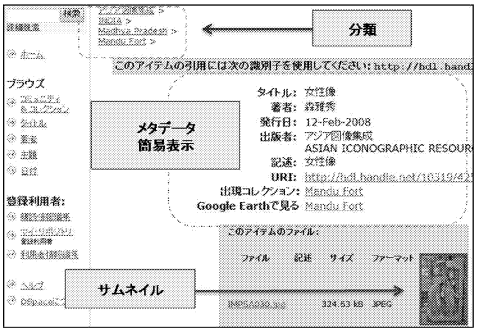


図 3　アジア図像集成

3.3.2 保守性の確保

(1)異種コンテンツの共存

異種コンテンツを同一リポジトリ上に共存させるには，コンテンツごとに適したメタデータ語彙を設定できることと，コンテンツごとに適した一覧表示を行える必要がある．これに対して，メタデータ語彙の設定や検索結果の一覧表示設定をコンテンツ種別に適切な設定に切り換えて利用できるようにコンテンツ別メタデータ語彙の設定法について考案した．図4に異種コンテンツを共存させた場合のイメージを示す．コンテンツの管理単位の最上位がルートコミュニティとなる木構造であることを利用して，種類の異なるコンテンツごとにルートコミュニティを分け，それぞれのコンテンツのルートコミュニティごとにメタデータ語彙の設定と一覧表示の設定を行えるようにした．

(2)コミュニティとコレクションの管理

図4からも分かるように，通常のリポジトリシステムでは，コミュニティとコレクションでアイテムを分類しコンテンツの管理を行っている．この管理のためにWebインターフェイスが準備されるのが一般的であるが，非文献コンテンツの場合，コンテンツの分類が複雑なものが多く，Webインターフェイスを介した方法では体系的な管理が煩雑になりがちである．これに対して，コンテンツ管理者が情報技術の専門家でなくても比較的なじみやすい，Excelファイルを用いたコミュニティとコレクション構造の管理法を考案した．図5は，アジア図像集成で用いたExcelファイルの記述とリポジトリ上での表示例である．

(3)一括登録

アイテムの登録においても，その数が数千件，数万件以上におよぶ非文献コンテンツに対して，Webインターフェイスで逐一登録する方法は，現実的とはいえない．これに対して，前項のコミュニティとコレクションの管理と同様に，コンテンツ管理者が所有のコンテンツを管理するために，情報技術の専門家でなくても比較的なじみやすいExcelなどの表計算ソフトを利用した方法を考案した．表計算ソフト上で管理されるメタデータの出力（TAB区切りまたはCSV形式）を読み込み一括登録が行える仕組みである．具体的には，ファイルの内容を読み込み，メタデータなどの解析を行い，システムが理解できる形式に変更しリポジトリに登録を行うスクリプトを作成した．図6にExcelなどの表計算ソフト上でのメタデータの記述形式を示す．

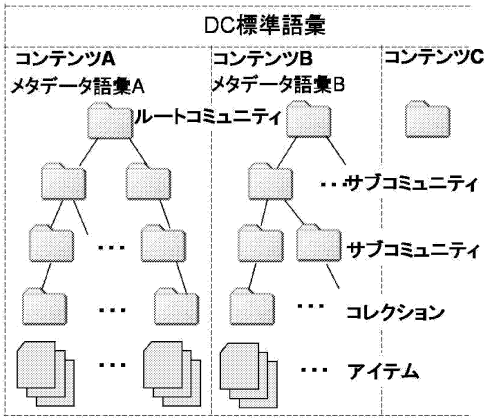


図 4 異種コンテンツを共存させた場合のイメージ

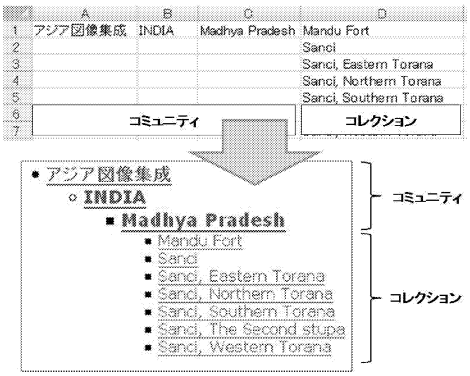


図 5　コミュニティとコレクションの記述例

3.3.3 Google Earthによる情報の可視化

非文献コンテンツでは，全文検索が利用できないため，コンテンツに付与されたメタデータのみによる検索が行なわれる．そのため，利用者は，文献コンテンツの検索と比べると，より適切な検索語句を指定する必要がある．一方，今回利用したアジア図像集成は，図像の出土地や所蔵地といった地理的な位置（地名）を示す情報を持っている．また，アジア図像集成に限らず，非文献コンテンツの中には，地理的な位置情報を持つコンテンツが少なくない．これらコンテンツが持つ地理的な位置情報を利用し，Google Earthと連携して，位置情報を地図上に表示することで可視化し，コンテンツの検索性の改善を図った．

具体的には，アジア図像集成はコレクション名が地名に相当するので，コレクションから座標を取得することとして，２つのスクリプトを作成した．一つ目は，登録されている地名から座標を取得し，Google Earth上に情報を表示させるKMLを生成するスクリプトを作成した．Google Earth上に表示させる情報には，地名の他，アジア図像集成へのリンクを含めることで，Google Earth上からアジア図像集成への検索も可能とした．もう一つは，与えられた地名から地図を表示するための視点を設定するKMLを生成するスクリプトを作成した．図7にGoogle Earthと連携した可視化システムの概要を示す．

3.3.4 他のリポジトリとの連携

他リポジトリとの連携が可能であることを実証するために，学内の学術情報を統一的に公開するポータルリポジトリを構築した．これは，学内に立ち上げられたリポジトリからメタデータをハーベスティングするハーベスタとしてリポジトリを構築し，個別のリポジトリにアクセスすることなく学内の学術情報を横断的に検索できるものである．

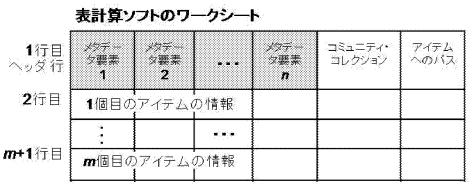


図 6　メタデータの記述形式

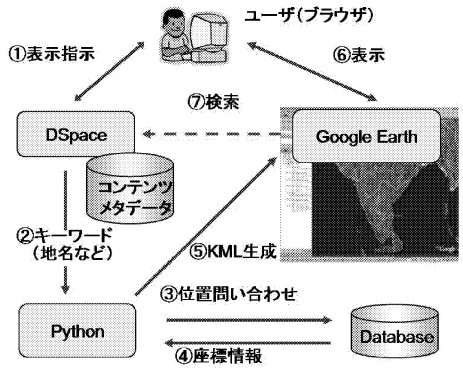


図 7 Google Earth と連携した可視化の概要

今回は，非文献コンテンツに最適化した本システムと文献コンテンツに適した当大学の図書館のKURA[10]で，非文献コンテンツと文献コンテンツを統一的に検索することができることを実証することができた．これにより，学術論文とそれに関係したデータ，図，写真などの非文献コンテンツの一括検索や各リポジトリに存在する異なる分野のコンテンツの横断的な検索が可能となる．

3.4 応用

本研究では，アジア図像集成の他に応用実装を行った．なお，応用実装として，前述のあけぼの衛星の観測データ，第四高等学校物理機器図録，およびe-Learning素材という特性の異なるコンテンツに対しても同一の手法を適用した．図8は，あけぼの衛星の観測データ，第四高等学校物理機器図録，およびe-Learning素材のそれぞれのアイテムの検索結果一覧画面である．

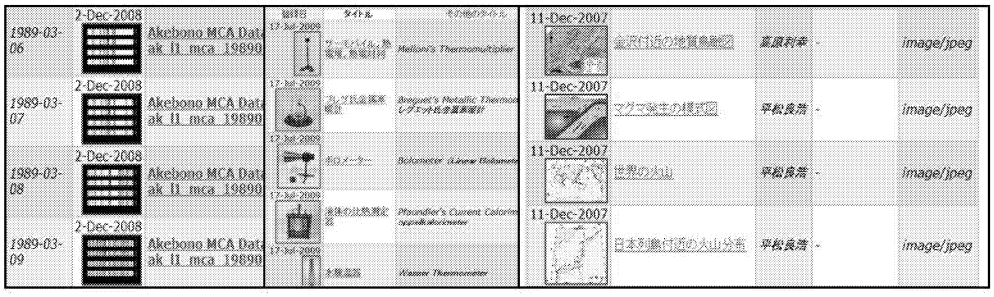


図 8　応用実装

【(a)：あけぼの衛星の観測データ，(b)：第四高等学校物理機器図，(c)：e-Learning 素材】

4 おわりに

非文献コンテンツをより広く公開し，教育・研究に有効利用されることを目的に，既存プラットフォームの問題点を改善し，可視性と保守性に優れた汎用性の高い学術情報リポジトリの構築を行った．

今回構築したシステムは，特定分野のコンテンツの特性に依存しない汎用的なものであり，全体目標として掲げた，大学における非文献コンテンツ公開のための共通プラットフォームとして広く応用可能である．また，状況に応じた柔軟な運用が可能で，コンテンツ管理者に掛かる負担も十分抑えることが可能と考えられ，蓄積している様々なコンテンツの公開が促進されると期待できる．

参考文献:

[1] 高田 良宏, 笠原 禎也, 他3名:「非文献コンテンツのための可視性と保守性に優れた学術情報リポジトリの構築」, 情報知識学会誌, Vol.19, No.3, pp.251-263, 2009.9.

[2] 高田 良宏, 笠原 禎也, 他2名:「多様なアクセス制限に対応した自然科学データベースシステムの開発」, 学術情報処理研究, No.11, pp.50-59, 2007.9.

[3] Y. Takata, Y. Kasahara, and T. Matsuhira: “Development of a Science Database System Applicable to Various Access Restrictions”, *Data Science Journal*, Vol.8, IGY32-IGY43, 2010.2.

[4] 高田 良宏, 笠原 禎也, 尾崎 友紀:「汎用データフォーマットを利用した自然科学データアーカイブスシステムの開発」, 学術情報処理研究, No.10, pp.5-14, 2006.9.

[5] Dublin Core Metadata Initiative, http://www.dublincore.org/ （2009.4.20参照）

[6] Open Archives Initiative: “Open Archives Initiative Protocol for Metadata Harvesting”, http://www.openarchives.org/pmh/ （2009.4.20 参照）

[7] DSpace Foundation: “DSpace.org”, http://www.dspace.org/ （2009.4.20 参照）

[8] 金沢大学:「アジア図像集成(Asian Iconographic Resources)」, http://wwwdb02.db.kanazawa-u.ac.jp/dspace, （2009.4.20 参照）

[9] 杉本 重雄:「メタデータについて：Dublin Core を中心として」, 情報知識学会誌, Vol.10, No.3, pp.53-58(2000).

[10] 金沢大学:「金沢大学学術情報リポジトリ」, http://dspace.lib.kanazawa-u.ac.jp/dspace/ （2009.4.20 参照）

部会報告

2010 年度第 3 回(通算第 12 回)情報知識学会関西部会研究会報告

河手太士 [1]
Futoshi KAWATE

田窪直規 [2]
Naoki TAKUBO

1 静岡文化芸術大学
Shizuoka University of Art and Culture

2 近畿大学
Kinki University

0 はじめに

これまで関西部会の研究会報告は、当学会サイトの関西部会のページで行ってきたが、今後は、国沢先生（学会誌編集長）の勧めもあり、学会誌で報告を行うことになった（ただし、この原稿を流用して、引き続き、当部会のページにも、報告記事を掲載していくつもりである)。

今回の研究会は、日本図書館研究会情報組織化研究グループとの共催であり、以下の報告は、同グループの河手太士氏によるものである。

1 研究会報告

日　時：3 月 13 日（土）14 時半～17 時
会　場：大阪市立浪速人権文化センター
テーマ：Net-based な情報流通の今における情報の組織化の機能理解
発表者：石川徹也氏（東京大学史料編纂所特任教授；筑波大学名誉教授)
共　催：日本図書館研究会情報組織化研究グループ
出　席：27 名

1.1 前提:「情報の組織化」目的理解

・世の中には多種多様にして大量の情報源が存在している。一方、人間は情報要求を持っており、その集合体が一つの情報要求マーケットとなる。
・情報要求マーケットと情報源をつなぐために、情報を目的別に取捨選択・収集し、整理・保存し提供しなければならない。これを「情報の組織化」とする。
・情報源が多種多様で大量であればあるほど、選択的収集技術や整理保存技術というものが発展しなければ、情報要求マーケットの要求に応えることができなくなる。

1.2 紙書籍を対象とする情報の組織化

・「電子書籍」という言葉に対して、これまでの紙媒体での書籍を「紙書籍」とする。
・今後は電子書籍の利用が隆盛期を迎える。
・図書館の特性に基づく収集方針に従って人手により収集が行われ、目録データなどが人手により作成される。非常にコスト高である。これが、現在の紙書籍の組織化の特徴である。
・普及期を迎える電子書籍に対する組織化はどのようにすればよいかが大きな悩みである。

1.3 電子書籍普及期における図書館の役割

・電子書籍の普及期においては、これから出版される書籍はデジタルコンテンツとして提供され、既刊本はデジタル化されて提供される。
・eBook や iPad、キンドルなどのユビキタス端末の普及により、電子書籍の利便性はますます向上する。
・紙書籍を読みたい場合には、プリント・オンデマンド端末を利用することにより、これまでの紙書籍（刊本）よりも利便性が高い利用が可能となった。
・刊本をデジタル化する技術や電子書籍に対するフルテキスト検索技術が確立されてきた。
・出版社が絶版本に商品価値があることに気づき、

long-tail 意識がめばえ、絶版本などの過去の資産をもう一度商品化しようとする意識が出版社に出てきた。

・long-tail 意識により刊本のデジタル化が促進され、さらに新たな本を出版する場合にも、紙で出版するよりもデジタルコンテンツとして提供されるということが急速に進むであろう。

・そうなったときに、従来の流通やプロバイダー、図書館はどのようなアイデンティティを確立し、どのような位置にいるべきであるかを考えることは大きな課題であると考える。

・図書館における内包的な問題として、インターネット検索と蔵書検索の違いや MARC データの限界、図書館利用の変化などがある。

・蔵書検索システムは図書館が開いていなければ調べることは出来ないが、インターネット検索だと可能である。蔵書検索は情報の入手までに物理的な制約が大きく、利便性においてインターネット検索との差は大きい。

・MARC-DB 検索は書名などの属性値データ検索や主題データ検索は可能であるが、情報やデータを検索することはできない。

・大学生の図書館利用ニーズが、「蔵書の利用」から「静かな空間であるから良い」「空調がきいていて快適である」「遅くまで利用できる」「インターネットが利用できる」などと大きく変化している。

・図書館は教育研究の場において、自己学習の場として無いと困る場所である。

・現在、インターネットでの情報については、その利用者も相当緊張して情報を利用しなければならない状況であるので、大学のコミュニティセンターを形成することにより「上質な知の共有」が可能となる。大学図書館は、大学のコミュニセンターにおける中心的な役割を担うことが必要であると考える。

1.4 電子書籍を対象とする組織化

・電子書籍を初めとするデジタルコンテンツを組織化することが今後は重要になってくる。

・電子書籍に対する収集・解析・保存・提供の機能は、テキスト処理システムが行う。

・収集方針に基づいてクローリングで自動的に収集され、書誌データなどは半自動で作成される。

・検索結果は、テキストの内容に基づく処理を行い TF-IDF などの数値を用いてランキング出力が可能となる。

・電子書籍は情報利用の効率化に寄与できるので、ますます電子書籍へシフトすることになる。

1.5 情報の組織化のための研究課題

・人間の情報要求に対して、現行の情報検索システムはほとんど対応できていない。

・人間の情報検索要求は、「事項知要求」「理由知要求」「五感知要求」に大きく分けることができる。

・「事項知要求」については、What, Who, Where, When の場合は、キーワードを入力すれば検索することができるが、Which, Why, How の場合は高機能化が必要であるが対応が可能。

・「理由知要求」は、状況判断が必要となり、状況判断を取り入れて情報を提供するということは、現時点においては実現できていないことから、システムでの実現も困難。

・「五感知要求」のうち、聴覚情報と視覚情報については計算機システムの援用により対応できるようになったが、臭覚情報・触覚情報・味覚情報についてはまだ対応できていない。

・これまでの情報組織化は紙書籍を管理するために行ってきた。これからは電子書籍の時代になり内容を適切に処理することが可能となってきたので、有効な情報（知識）を提供することが情報の組織化ではないかと考える。

1.6 さいごに

・情報要求は一つのマーケットであり、それをつなぐのが情報の組織化である。

・情報の組織化は、紙の組織化から情報の組織化が現在行われている段階である。次のステップとしては知の組織化、知の発見組織化となる。

・知の組織化や知の発見組織化については研究に着手されたばかりではあるが、われわれはそのような流れの中にいることを自覚し研究をすすめていくことが必要だ。

第 15 回情報知識学フォーラム
「多様化する電子書籍端末と学術情報流通」

情報知識学フォーラム実行委員会

iPad, Kindle は電子書籍のブームを起こしましたが、現状はむしろ多様化しているように見えます。そこで標記のテーマを主に技術的観点から考えるため、次のとおり、大学、出版、印刷の分野から専門家を招き、講演と総合討論を行います。皆様、どうぞご参加ください。

日　時：2010 年 12 月 4 日(土) 13:00～17:20
会　場：慶應義塾大学三田キャンパス第 1 校舎 111 教室（東京都港区三田 2-15-45）
主　催：情報知識学会
後　援：(社)情報科学技術協会　**協　賛**：(社) 日本印刷学会

プログラム

13:00-13:10 開会挨拶　根岸正光（情報知識学会会長）
13:10-15:40 講演
講演 1: 「**書籍の電子化がもたらすもの ―素朴な疑問と素朴な期待―**」
　　杉本重雄（筑波大学教授）
講演 2: 「**電子書籍交換フォーマットの現状と標準化**」
　　植村八潮（東京電機大学出版局 局長）
講演 3: 「**EPUB の多国語対応に向けた取組と事例報告**」
　　秋元良仁（凸版印刷(株)情報技術研究室 シニア研究員）
－－－　休憩　（10 分）　－－－
15:50-17:10 講演および総合討論
講演 4: 「**電子書籍フォーマットの研究動向と学術情報流通への課題**」
　　原田隆史（慶應義塾大学准教授）
総合討論:
　司会: 原田隆史
　パネリスト: 杉本重雄、植村八潮、秋元良仁
17:10-17:20 閉会挨拶　石塚英弘（フォーラム実行委員長）
17:30- 懇親会

参加費と事前登録

・会員（後援／協賛の学協会会員も含む）：無料
・非会員：論文集代 3000 円(一般)／1500 円(学生)
・情報知識学会への当日入会可
・事前申込フォームを学会サイトにて 10 月末開設

お問い合わせ先

情報知識学会事務局
- 〒110-8560 東京都台東区台東 1-5-1（凸版印刷(株)内）
- E-mail: jsik(at)nifty.com　URL: http://wwwsoc.nii.ac.jp/jsik/

事務局からのお知らせ

［1］年会費未納のかたは納入をお願いします

平成 22 年度は、本年 4 月 1 日から来年 3 月 31 日までの 1 年間です。年会費未納のかたへ本年 8 月下旬に請求書を郵送しました。早急に下記へお振り込みください。

退会する場合、年会費未納のかたには本年 4 月 1 日から退会届提出日までの年会費を四半期単位の割引で計算し、事務局から改めてお知らせいたしますので、その金額を下記へお振り込みください。

1 年間の年会費は正会員８千円、学生会員・ユース会員・シニア会員は４千円です。過去数年分未納のかたは合計額を納入してください。特定の請求書が必要な場合、その旨を事務局へ電子メールその他でお知らせくだされば郵送いたします。

１．振込先（振込手数料はご本人負担でお願いします）

ａ．郵便振替口座 00150-8-706543 情報知識学会（代表　根岸正光）
ｂ．ゆうちょ銀行　〇一九店(ゼロイチキュウ店）当座 0706543 情報知識学会
（代表　根岸正光）

２．納入した年月日の確認方法

お手元へ届く学会誌などの宛名ラベルをご覧ください。［　］内に過去４年間、ご自分の納入日が印字されているので確認できます。　納入年（西暦の下２桁）、月（２桁）、日（２桁）の６桁です。年会費を滞納している場合は、［未納］と表示してあります。金融機関へ振り込まれた日から、事務局へ通知が届き、宛名ラベルに印字、発送するまで約 10 日かかりますので、ご了承ください。

［2］電子メールアドレスをお知らせください

毎月発行している情報知識学会メールマガジンは会員の皆様全員に読んでいただきたい内容です。電子メールアドレスを未登録のかたや最近更新されたかたは、ぜひ事務局 jsik@nifty.com へご連絡ください。

［3］電話でのお問い合わせ

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

入会ご希望のかたには入会申込書を、郵送またはＦＡＸ送信でお届します。当学会のホームページ http://www.jsik.jp/から直接申し込むこともできます。

情報知識学会事務局
〒110-8560　東京都台東区台東 1-5 凸版印刷㈱内
TEL:03-3835-5692　　FAX:03-3837-0368
E-mail:jsik@nifty.com　　URL:http://www.jsik.jp

情報知識学会誌 Vol. 20, No. 3　2010 年 10 月 22 日発行 編集・発行情報知識学会

頒布価格 3000 円

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

会長　根岸　正光

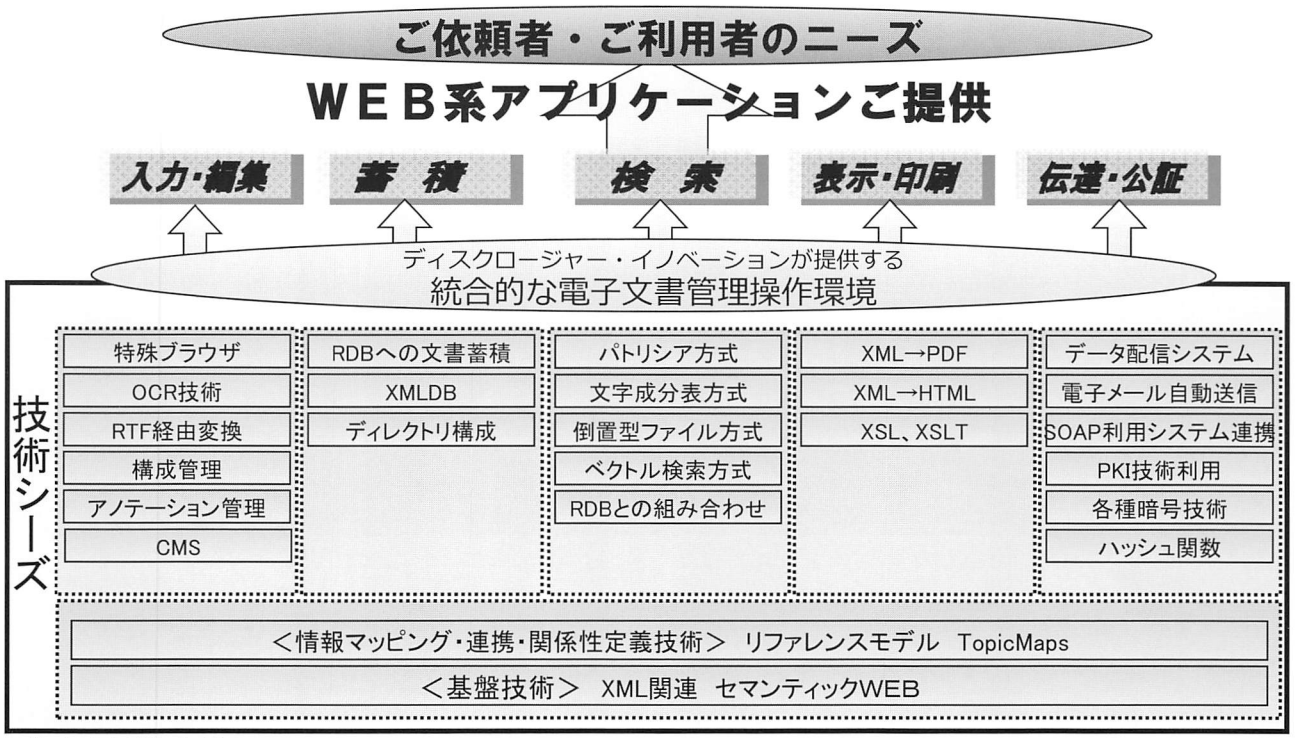
事務局　〒 110-8560　東京都台東区台東１－５－１　凸版印刷（株）内

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

URL: http://www.jsik.jp

Disclosure Innovation Inc.

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



主な事業内容

情報提供サービス機構の実現

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

電子文書を利用した社会機構の調査

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

電子文書関連ツール＆サービス開発と販売

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

ネットワーク関連(管理と開発)

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

文書情報管理

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

ディスクロージャー・イノベーション株式会社
シナジー・インキュベート事業部
171-0033 東京都豊島区高田三丁目23番10号 宝印刷本社別館(5号館クリスタルエイトビル) 5F.
Tel : 03-5985-0920 Fax : 03-5985-0921 URL : http://www.di-inc.co.jp/

Journal of Japan Society of Information and Knowledge

Contents

情報知識学会誌 第20巻3号 2010年10月22日発行
編集兼発行人 情報知識学会 〒110-8560 東京都台東区台東1-5-1 凸版印刷㈱内
TEL:03(3835)5692 FAX:03(3837)0368 E-mail:jsik@nifty.com
URL: http://www.jsik.jp/ (振替：00150-8-706543)

学術刊行物 ISSN0917-1436