

# 目 次

## まえがき

1. 「情報の巨大集積化と利活用基盤技術開発」連携施策群の意義と成果の概要	1
2. 「情報の巨大集積化と利活用基盤技術開発」連携施策群の成果報告	7
3. 連携施策群の活動報告	
3.1 情報大航海プロジェクト：経済産業省	19
3.2 電気通信サービスにおける情報信憑性検証技術に関する研究開発：総務省	29
3.3 革新的実行原理に基づく超高性能データベース基盤ソフトウェアの開発：文部科学省	39
3.4 多メディア Web 解析基盤の構築及び社会分析ソフトウェアの開発：文部科学省	43
3.5 センサ情報の社会利用のためのコンテンツ化（補完的課題）：京都大学（研究代表機関）	45
4. 連携施策群の基盤技術	
4.1 情報の巨大集積化と利活用基盤技術開発（技術連携マップ）	55
4.2 技術連携マップの基盤技術説明	
4.2.1 情報大航海プロジェクト	56
4.2.2 情報信憑性検証技術	82
4.2.3 超高性能データベース基盤ソフトウェア	86
4.2.4 センシング Web	87
4.3 基盤技術の活用例	
4.3.1 行動情報に基づいた最適な情報やサービスの提供	90
4.3.2 高度な予防型安全管理システムの実現	91
4.3.3 生活習慣に関する情報等に基づいた「情報薬」の提供	92
4.3.4 ブログランキングサービス	93
4.3.5 センシング Web	95
5. 連携施策群の情報発信活動紹介	97

おわりに

## まえがき

科学技術連携施策群「情報の巨大集積化と利活用基盤技術開発」  
コーディネーター／主監 西尾 章治郎

本成果報告書は、平成 19 年度から 21 年度まで 3 年間にわたり活動を行ってきた総合科学技術会議科学技術連携施策群「情報の巨大集積化と利活用基盤技術開発」における多くの成果を集大成したものです。

近年、電子化されたデータは爆発とも言える状況でインターネット上に溢れています。その量は、平成 22 年に 988 エキサバイト（10 の 18 乗バイト）に達すると言われています。これらのデータは知の泉であり、我々はそこから知を紡ぎ、社会において多様に有効利用することがますます重要になっています。しかし、現状では、二つの大きな課題があります。一つは、情報の量の問題です。現在、直接検索可能な Web 情報は全体の 0.1% 未満であり、95% の情報が整理・構造化されていません。これらをうまく活用すればもっと有意義な「価値」の創造、例えば、新たなビジネス創出が期待できます。もう一つは、情報の質の問題です。現状では、Web の情報に対し、ユーザサイドからその信憑性を確かめる手立てがありません。一般に、自分の欲しいデータへアクセスする場合、検索エンジンが広く利用されています。しかし、現在の検索エンジンは、ゴミ情報や有害な情報を上位ランクに提示する可能性などの問題点があり、弊害も多いことが指摘されています。

「情報の巨大集積化と利活用基盤技術開発」の目的は、これらの課題を解決し、安心安全な社会を実現するために府省連携を推進し、次世代の知的情報利活用基盤を構築することです。具体的には、文部科学省の「革新的実行原理に基づく超高性能データベース基盤ソフトウェアの開発」、経済産業省の「情報大航海プロジェクト」、総務省の「電気通信サービスにおける情報信憑性検証技術に関する研究開発」、これらを補完する「センサ情報の社会利用のためのコンテンツ化」の四つの施策を連携して開始しました。これらの施策は、実施期間が異なっているため、連携施策群としては一番短い 3 年に基準を合わせ、平成 21 年度をもって終了します。この 3 年間で当初の目的以上の大きな成果が得られました。例えば、新たな施策として文部科学省の「多メディア Web 解析基盤の構築及び社会分析ソフトウェアの開発」を加えることができたことは、連携施策群の重要性が十分に認知されてきた証左と言えます。また、技術だけでなく法制度面にも考慮し、「情報大航海プロジェクト」が著作権法改正の成立に少なからず寄与したことは大きな成果と言えます。今後ともますます重要になると思われる科学技術イノベーションというキーワードは、まさに技術的なイノベーションだけでなく、法制度の改革を始めとする社会的なイノベーションをも包括しています。本連携施策群はその先駆け的な意味を果たせたと確信しています。

今後、このような国家プロジェクトをさらに継続する機会を設け、国際競争力のさらなる向上、情報の巨大集積化と利活用に対し日本がどのように貢献できるかが大きな課題です。すなわち、平成 21 年度で終了する 2 施策と、継続する 3 施策に関しては、何らかの形で議論の場を設け、最後まで有機的な連携を取っていくことが重要であると考えています。

末筆ながら、本報告書が、関連研究者はもとより、一般の方々にも広く読まれ、新しい研究開発やビジネス創出を誘発する機会の拡大に寄与できたとしたら望外の喜びです。

## 1. 「情報の巨大集積化と利活用基盤技術開発」連携施策群の 意義と成果の概要

# 1. 「情報の巨大集積化と利活用基盤技術開発」 連携施策群の意義と成果の概要

科学技術連携施策群「情報の巨大集積化と利活用基盤技術開発」  
コーディネーター／主監 西尾 章治郎

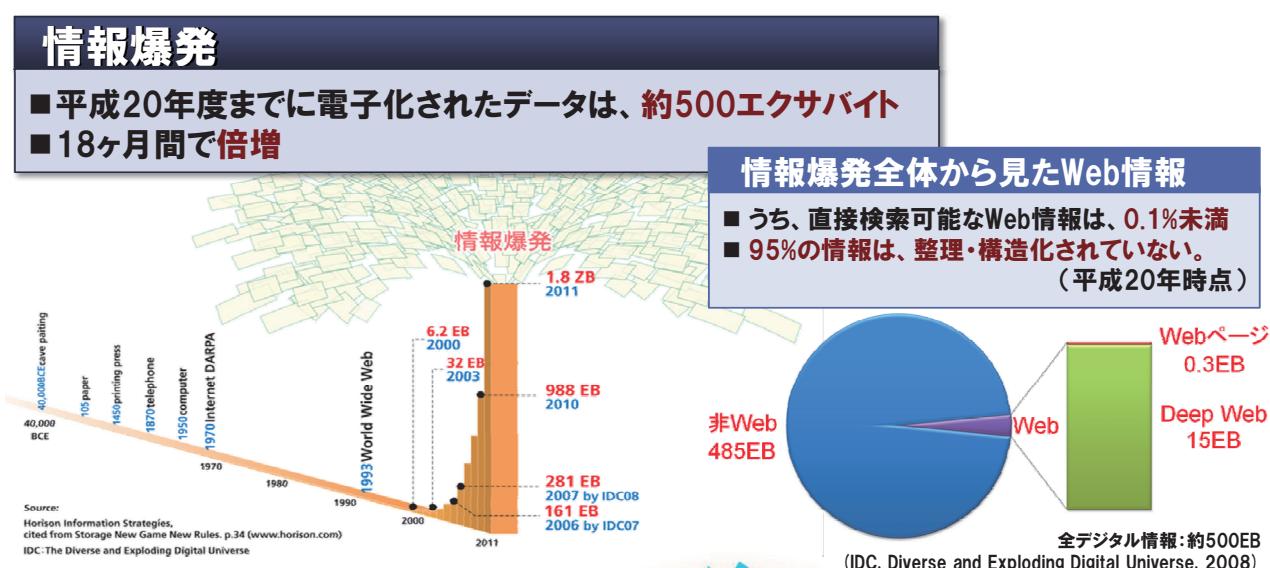
## ● 背景：

### ◆ 情報爆発時代の到来

- デジタル情報は指数関数的に増大(18ヶ月間で倍増)
  - 平成20年度までに電子化されたデータは、約500エクサバイト  
(1エクサバイト=10億ギガバイト、DVDメディア約2億枚に相当)

しかし、情報が増えるほど、  
本当に欲しい情報は埋もれてしまう。

- 直接検索可能なWeb情報は全体の0.1%未満  
(平成20年時点)
- ごみ情報や有害情報が含まれている。
  - 従来検索では上位ランクに提示されてしまう恐れがある。



### 情報爆発現象の捉え方

- 世の中には、これほど沢山の情報があるので、うまく活用すればもっと有意義な「価値」が榨り出せる。
- 新たなビジネスチャンス

1EB=1エクサバイト=10億ギガバイト  
1ZB=1ゼッタバイト=1,000エクサバイト



## ● 背景(続):

もし、扱える情報の量が増やせれば、  
ごみ／有害情報を除くことができれば  
➤ 新たなビジネスチャンスにつながる。

そのためには、

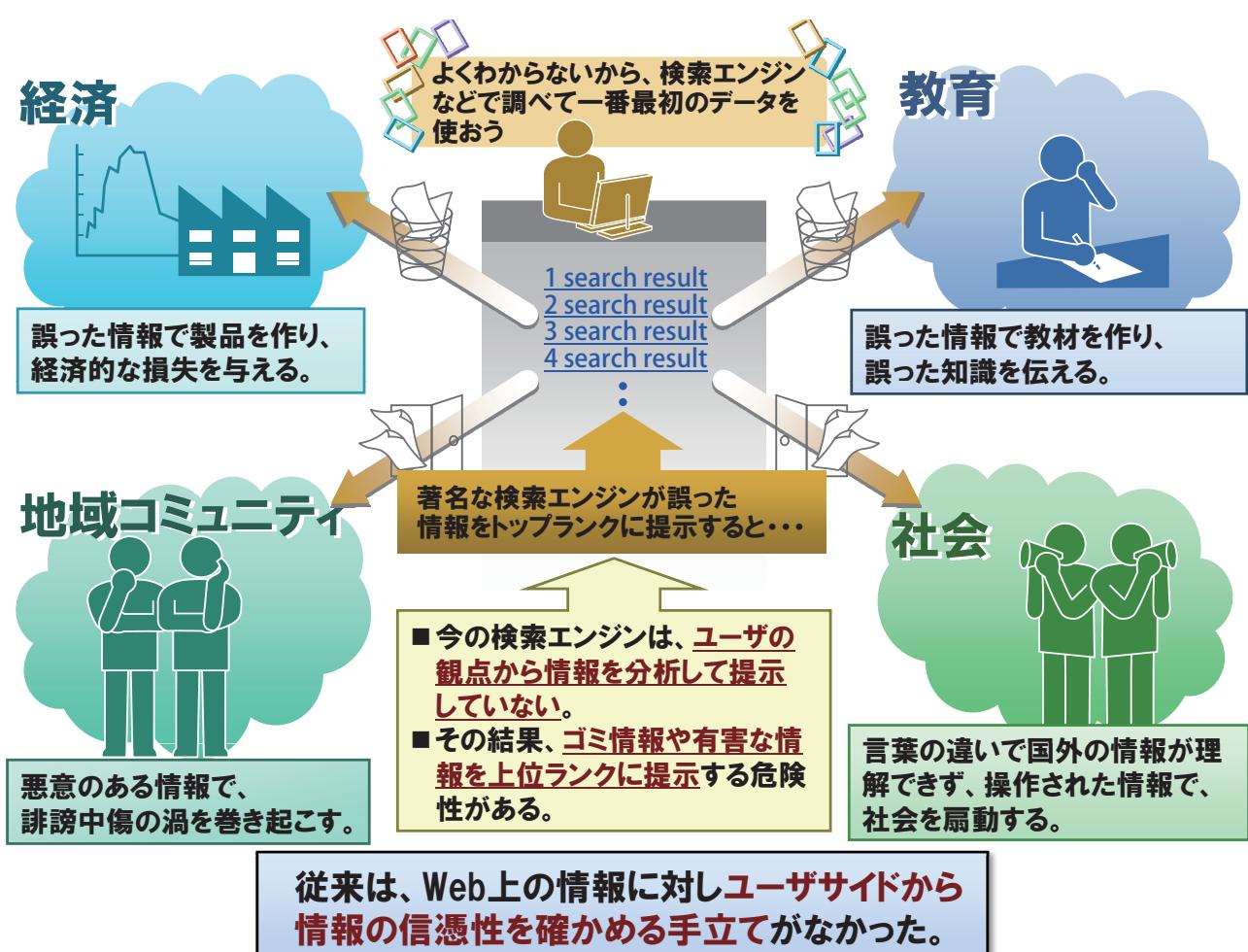
**情報の量と質の問題への対処が必要**

### 1. 情報の量の問題への対処

- 課題: 95%の情報は整理、構造化されていない。
  - 整理、構造化する手立てを確立し、誰もがアクセス可能に
- 課題: データへの検索スピードが遅い。
  - 高速化により、瞬時により多くの情報へのアクセスが可能に

### 2. 情報の質の問題への対処

- 課題: 情報の信憑性を確かめる手立てがない。
  - 信憑性を検証する手立てを確立し、ごみ／有害情報を排除

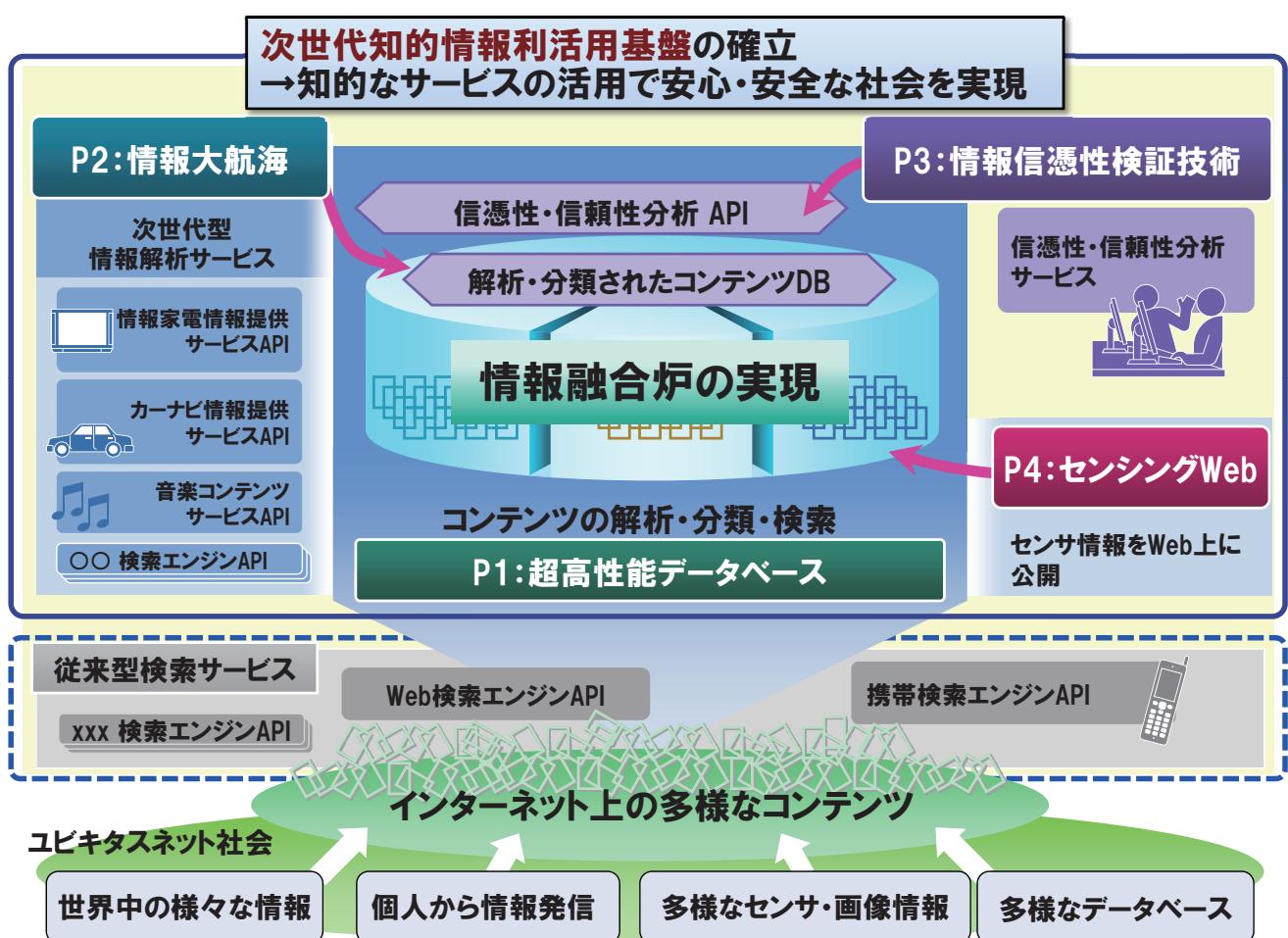


## ●目標:

独自の情報サービスを提供するためにあらゆる情報(コンテンツ)を簡便、的確、かつ安心して収集、解析、管理する**次世代の知的な情報利活用**のための基盤技術を開発する。

## ●意義と実現方法:

- 意義:既存の技術では管理することが困難な大量の情報の中から、**信憑性**を判断でき有益な情報を高速に見つけ出すことを可能とし、様々な情報サービスの基盤となる**次世代知的情報利活用基盤**を構築する。
- 実現方法:具体的には、データを管理する**超高性能データベース**(プロジェクト1:P1)、情報を活用するサービス技術(プロジェクト2:P2)、得られた**情報の信憑性**を解析する**情報分析技術**(プロジェクト3:P3)、さらに非Web情報であるセンサ情報の活用技術(プロジェクト4:P4)を融合させる。



## ● 主な成果(3年間の総括):

### 1. 施策面:当初計画以上の成果を達成

#### ①各施策:当初目標値を大幅に上回る成果を実現

超高性能データベース基盤では、当初目標値を上回る従来比20倍の高速化を実現 → 現実的な時間で検索可能な情報量が20倍に

#### ②適用領域拡大:新規施策の追加、最先端研究開発支援プログラムへの一本化などにより適用領域を拡大 → より多様、より多くの情報を対象に

### 2. 法制度面:情報利活用する際の法制度面整備に大きく寄与

#### ①現行法:著作権法の一部を改正する法律成立(平成21年6月)

情報検索サービスを実施するための複製が可能に

#### ②今後:次世代パーソナルサービス推進コンソーシアムを設立(平成21年11月)。継続的に検討する体制が確立された。

### 3. 情報発信:国内はもとより国際的にも積極的に活動

#### ①シンポジウムでは非常に高い評価(アンケート結果)

#### ②主要会議(国内、国際)等で情報発信

#### ③ホームページの活用により成果を国内外に発信

#### (1) 施策

##### 当初計画

- 1. 情報大航海プロジェクト
- 2. 情報信憑性検証技術
- 3. 超高性能データベース基盤ソフトウェア
- 4. 補完的課題:センシングWeb

##### 主な成果

###### ■各施策:当初計画以上の成果を達成

超高性能データベース基盤では、当初目標値を上回る従来比20倍の高速化を実現

###### ■適用領域の拡大:

- ・新規施策を追加:急激に増加しているビデオデータなどのリッチデータを有効利用するために「Web社会分析プロジェクト」を追加
- ・最先端研究開発支援プログラムへ一本化:我が国のソフトウェア産業における国際競争力強化の観点から「超高性能データベース基盤ソフトウェア」を同プログラムへ発展的に一本化(平成22年度)

#### (2) 法制度面

##### 当初計画

情報利活用する際の法制度に関わる諸問題の検討

##### 主な成果

###### ■著作権法の改正への働きかけを実施し、著作権法の一部を改正する法律成立(平成21年6月)

産学主導でパーソナル情報のサービス適用を検討する「次世代パーソナルサービス推進コンソーシアム」の設立(平成21年11月)

#### (3) 情報発信

##### 当初計画

- シンポジウム開催
- 国内、国際会議
- ホームページ活用

##### 主な成果

###### ■シンポジウム、国内、国際会議:積極的に発信

###### ■ホームページによる成果の国内外発信

欧州において連携群と同様なプロジェクトを推進しているクエロ(フランス)、ファロス(イタリア、ノルウェー等9ヶ国が参画)からの招待講演

## ● 主な成果(従来技術との比較):

### 1. 量的側面:巨大集積化によりこれまで扱えなかった量の情報をリアルタイムに処理可能とし、新たなイノベーションを実現

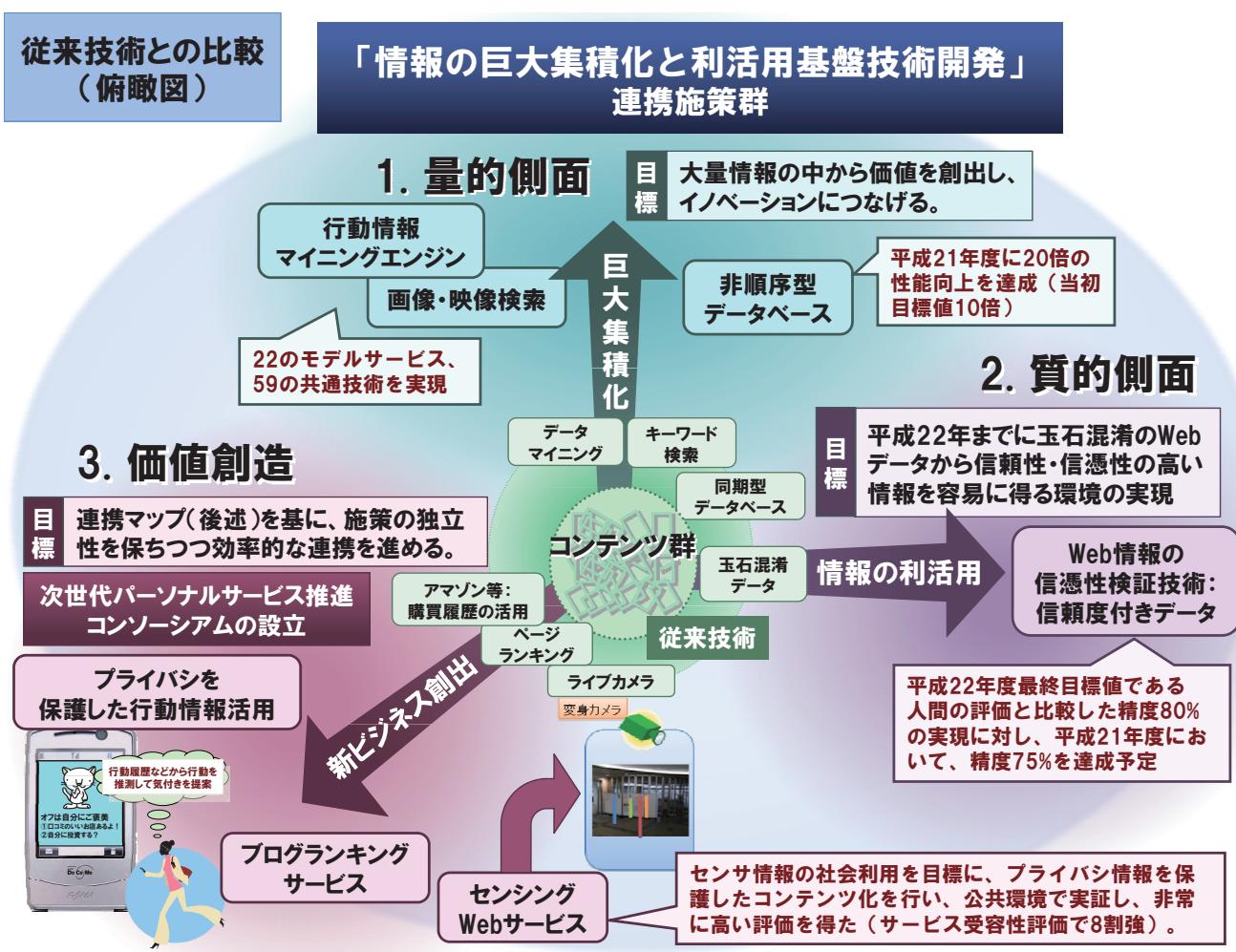
- 従来のキーワード検索より情報量が多くユーザビリティの高い「画像・映像検索技術」、移動ログ情報から個人の行動特性や場の特性を見出す「行動情報マイニングエンジン」など、従来、整理・構造化されていなかった非Webなどの領域を扱うのに必要な**59の共通技術**を開発し、**新たなイノベーション**として**22のモデルサービス**を推進した。

### 2. 質的側面:従来技術に対し、具体的な情報の信頼度を付加した「Web情報の信憑性検証技術」など着実に進捗

- そもそも従来はこのような尺度で信憑性検証はなされていなかった。
- 人間の評価に対し、**精度75%**に迫る。

### 3. 価値創造:新ビジネス創出を視野に入れた実証実験を推進

- センサ情報を誰でも閲覧・利用できる「センシングWebサービス」では、**高い評価(受容性評価で8割強)**を得た。
- 産学主導による「次世代パーソナルサービス推進コンソーシアム」が設立され、パーソナルサービス分野の進展に大きく寄与した。



## ● 主な成果(連携成果)と今後の展開

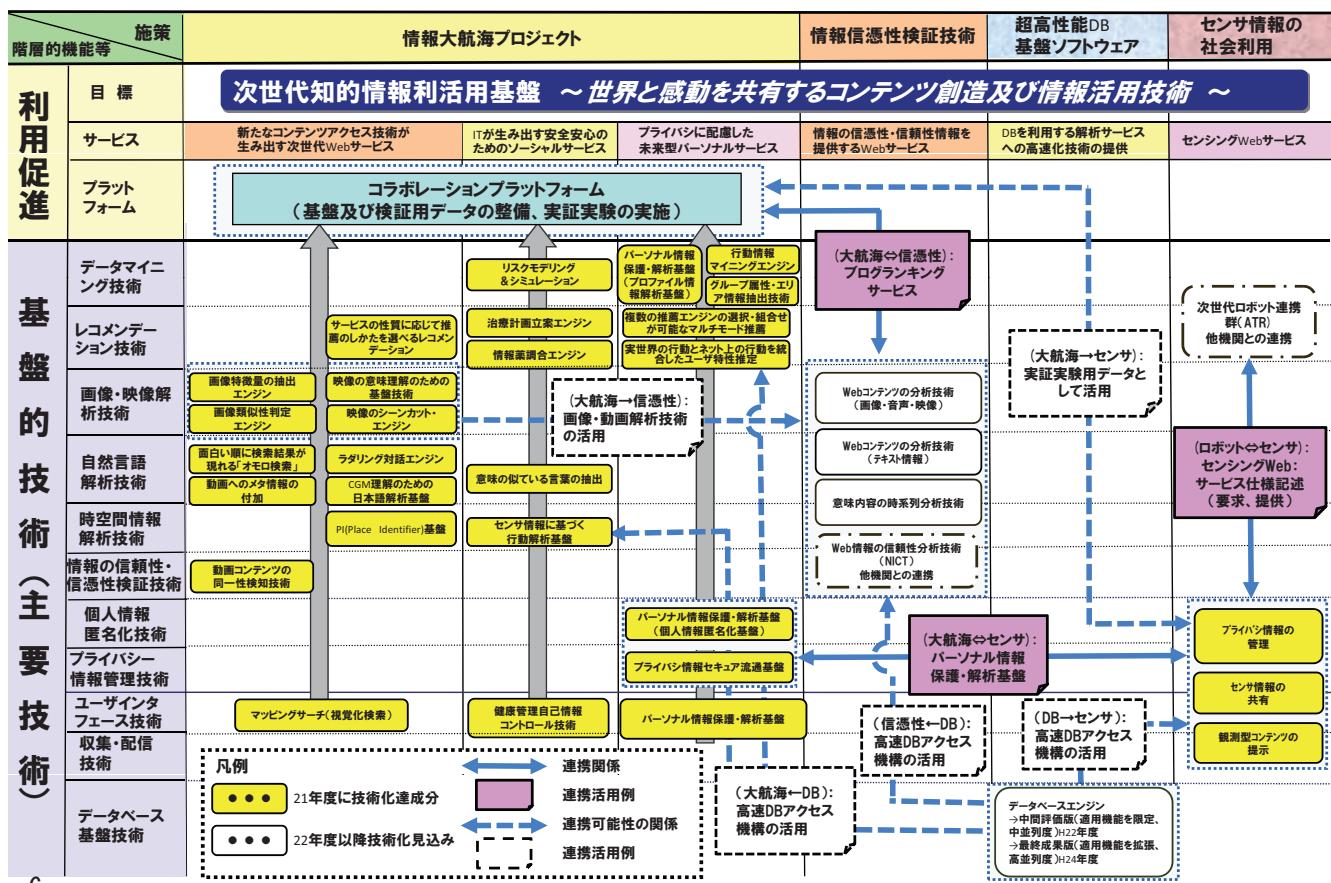
### 1. 連携成果: 主な技術と連携成果を俯瞰した連携マップ作成

- ① 主な技術: 連携マップにより各施策間の連携シナリオを想定しながら開発を進め、「基盤的技術」、「実証サービス」などを実現
- ② 連携成果: 各施策の連携を強力に推進し、新たな技術、サービスを創出(下図の連携活用例)

### 2. 今後の展開: 「情報の巨大集積化と利活用」は重要な課題

- ① 「情報の巨大集積化と利活用」における成果をさらに広く活用可能とするために、例えば、パブリッククラウドといった産官学クラウドの基盤構築などを提言
- ② 今後の持続的な研究開発を推進するにあたり関連諸機関の連携体制の一層の強化を図り、国際競争力のさらなる向上を目指す。

### 情報の巨大集積化と利活用基盤技術開発連携群(連携マップ)



2. 「情報の巨大集積化と利活用基盤技術開発」連携施策群の  
成 果 報 告

## 2. 「情報の巨大集積化と利活用基盤技術開発」連携施策群の成果報告

### 科学技術連携施策群「情報の巨大集積化と利活用基盤技術開発」 コーディネーター／主監 西尾 章治郎

#### 概要

総合科学技術会議（内閣府）のイニシアティブのもと、経済産業省、総務省、文部科学省は、平成19年度より「情報の巨大集積化と利活用基盤技術開発」連携施策群を推進してきた。各省のプロジェクトは、3～5年間と期間は異なっているものの、日々刻々爆発的に増加しているインターネット内の巨大データを集積し利活用できる知的情報アクセス基盤の構築を目指している。今回、連携群としては、最終年度にあたることから、これまでの3年間に活動した内容の総括、具体的には、連携の観点から各施策において新たに開発した成果に加え法制度面や情報発信による成果についても総括した結果について述べる。

#### 1. 背景

国連貿易開発会議（UNCTAD）による平成18年度情報経済報告によると、世界のインターネット利用者は平成17年度に前年度比19.5%増の10億2000万人を超えており、初めて10億人の大台を超えたことを報じている。また、ブログやソーシャルネットワークサービス（SNS）など、誰でも全世界に情報発信できる新たなメディアの出現によって、人間、さらにセンサやマシンによって創出される情報量は、急激に増大している。図1に示すように、米国の市場調査会社IDCが平成19年3月に発表した”The Expanding Digital Universe”と題する白書によれば、全世界で電子データとして創出あるいは獲得され、蓄積されている情報量は、平成18年に161エクサバイト（1エクサ（Exa）バイト= $10^6$ 兆バイト）であったのに対して、4年後の2010年には約6倍の988エクサバイトという想像を絶する量に達すること、つまり毎年約1.5倍の割合で増大し続けることを予測している。

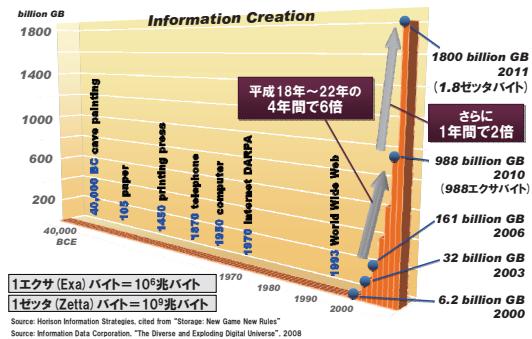


図1 情報爆発時代の到来

先の調査データが示すように、情報の巨大集積化時代の到来が現実のものとなっている現在、学術分野のみならず、日常生活に密着した環境やビジネスの分野も、非常に深刻な問題に直面している。つまり、情報発信量の著しい増加に対し、一人の人間が処理できる能力には限界があるという現実である。電子メールにしても、一定量のメール受発信であれば極めて効率的であるが、処理に数時間をするほど受信量が増大すると、かえって知的生産活動を阻害しかねない。人々が、必要な情報を探し出すために知的活動のうちの多くの時間を割かざるを得ないという状況は、

最近の統計データでも示されており、巨大な情報空間からの確に必要な情報を引き出す技術が重要になってきている。そのような技術を実現するためには、巨大集積された情報を整理し、引き出しやすい索引（インデックス）などを付与することで、必要な情報を検索しやすくするなど、利活用のための基盤技術が重要になる。つまり、容易に知を紡ぐことが可能な情報プラットフォームの構築が求められている。

情報の巨大集積化時代を迎えるに当たり、もう一つ大きな問題が浮上している。蓄積されている情報そのものの信頼性の問題である。いくら大量に情報が集積されたとしても、ごみ情報や有害な情報が多分に含まれているのではユーザは安心できない。しかし、それらを完全に除外することは難しく、現実的にどのように対応していくかは深刻な課題である。一般に、欲しい情報あるいはデータへのアクセスの手立てとして検索エンジンを利用しているが、従来の検索エンジンでは、ユーザの観点から情報を分析しているとは言い難い。例えば、ごみ情報や有害な情報を上位ランクに提示してしまうなど、いろいろな問題点、弊害が数多く見受けられる。

## 2. 目的

このような情報の巨大集積化時代における諸課題を解消すべく、「情報の巨大集積化と利活用基盤技術開発」連携群は、平成 19 年度より開始された。「情報の巨大集積化と利活用基盤技術開発」の各施策は、総合科学技術会議における戦略重点科学技術「世界と感動を共有するコンテンツ創造及び情報活用技術」に属しており、その目標とするところは、図 2 に示すように、独自の情報サービスを提供するためにあらゆる情報（コンテン

ツ）を簡便、的確、かつ安心して収集、解析、管理する次世代の知的な情報利活用のための基盤技術を開発するものである。

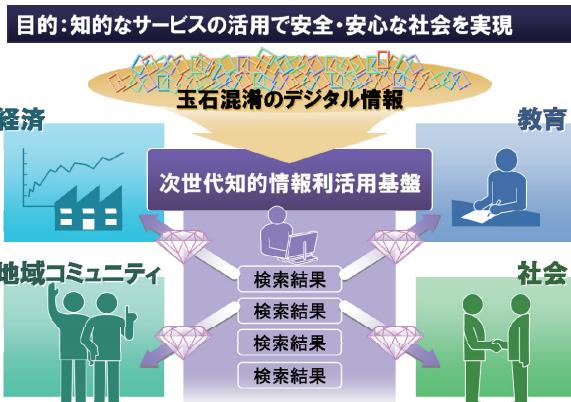


図 2 知的情報アクセス技術基盤と社会

すなわち、既存の技術では管理することが困難な大量の情報の中から信憑性が判断できる有益な情報を高速に見つけ出すことが可能となる。

これにより、例えば、データを管理するデータベース、情報を活用するサービス技術、得られた情報の信憑性を解析する情報分析技術を融合させることで、次世代の利活用基盤の構築が可能となる。

## 3. 推進体制

### a. 科学技術連携施策群

従来の省庁間の縦割り的な科学技術プロジェクトの推進を改め、重複排除し、特に、内容的に連携できるものに関しては、優先的にあるいは重点的に取り組むという観点より、総合科学技術会議のリーダーシップの下、推進している施策に連携施策群がある。平成 21 年度現在、図 3 に示すように、六つのプログラムが活動している。「情報の巨大集積化と利活用基盤技術開発」連携群は、情報通信関係では、唯一のものとして活動している。

プロジェクトチーム	科学技術連携施策群
ライフサイエンスPT	食料・生物生産研究
	臨床研究・臨床への橋渡し研究
情報通信PT	情報の巨大集積化と利活用基盤技術開発
環境PT	総合的リスク評価による化学物質の安全管理・活用のための研究開発
ナノテクノロジー・材料PT	ナノテクノロジーの研究開発推進と社会受容に関する基盤開発
社会基盤PT	テロ対策のための研究開発

参考:<http://www8.cao.go.jp/cstp/project/main.html>

図3 科学技術連携施策群の推進体制

b. 「情報の巨大集積化と利活用基盤技術開発」連携群

(1)連携体制

「情報の巨大集積化と利活用基盤技術開発」連携群は、図4に示すように、文部科学省の「革新的実行原理に基づく超高性能データベース基盤の開発」、経済産業省の「情報大航海プロジェクト」、総務省の「電気通信サービスにおける情報信憑性検証技術に関する研究開発」、さらにこれらを補完する課題として、非Webにポイントを置いた「センサ情報の社会利用のためのコンテンツ化」という四つのプロジェクトを連携して平成19年度から開始した。

図4に示すように、各々のプロジェクトは、そのプロジェクトの期間が異なっており、一番短いものは3年である。これに基準を合わ

せると、平成21年度がこれらの四つのプロジェクトがすべて動いている最終年度であることから、平成21年度をもって活動の一区切りとして、連携施策群を終了する。

(2)運営体制

「情報の巨大集積化と利活用基盤技術開発」連携群の運営体制は、独立行政法人科学技術振興機構の支援により、以下の体制で運営している。

- ・ コーディネーター/主監：西尾 章治郎
- ・ タスクフォース委員  
木俵 豊（情報通信研究機構）  
喜連川 優（東京大学）  
東倉 洋一（国立情報学研究所）  
松山 隆司（京都大学）  
安田 浩（東京電機大学）  
横田 治夫（東京工業大学）  
鷲尾 隆（大阪大学）  
岩爪 道昭（情報通信研究機構）
- ・ 総合科学技術会議事務局  
大石 善啓（内閣府）  
金谷 学（内閣府）  
内元 清貴（内閣府）
- ・ 科学技術連携施策群支援業務室  
谷本 茂明  
林 敬三郎

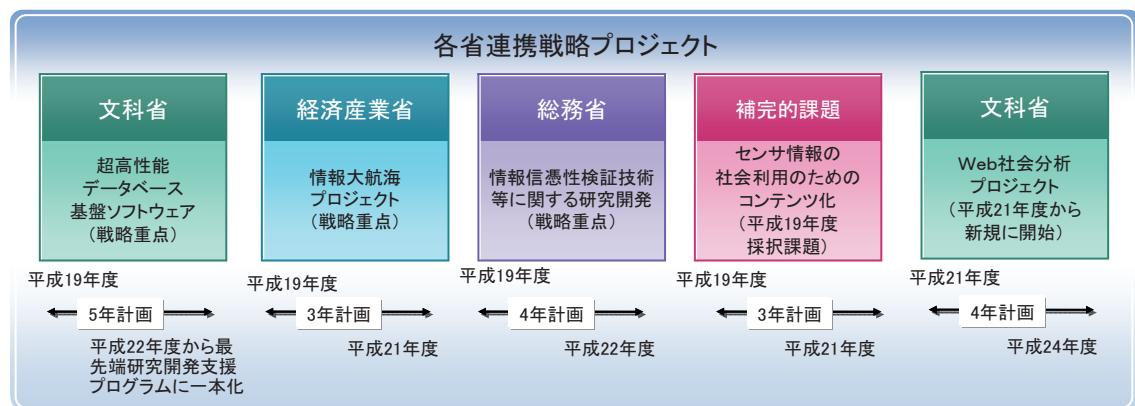


図4 「情報の巨大集積化と利活用基盤技術開発」連携群の連携体制

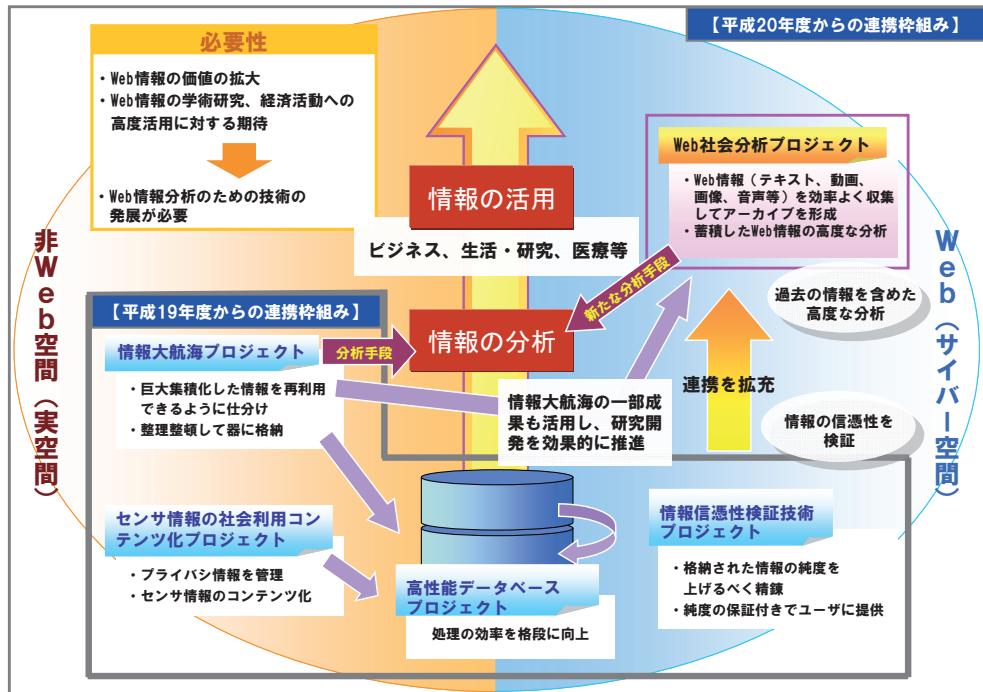


図5 情報の巨大集積化と利活用基盤技術開発連携群の概要

#### 4. 成果

「情報の巨大集積化と利活用基盤技術開発」に関わる連携施策群の具体的な内容として、図5を用いて説明する。

図5に示すように、最初に、巨大集積化した情報を再利用できるように仕分けを行い、これを整理整頓して棚に格納する。その超大規模なデータを共用できる形でデータベース化し、その検索機能の構築、さらに、これらの基盤上で色々なサービスを提供する。これが情報大航海プロジェクトである。次に、このようなデータがいくら蓄えられたとしても、それが信頼性を有していないと、有効利用できない。従って、情報の信憑性検証技術プロジェクトがある。さらにユビキタスやアンビエント化に伴い、センサ情報の活用が今後、重要となってくる。このセンサ情報をプライバシに配慮してコンテンツ化する補完的プロジェクトがある。このようなデータが大量にあった時に、データベースの性能も重要なことから、超高性能データベースプロジェクトがある。

さらにここでは非 Web 空間、あるいは Web 空間を対象にしているが、現在、Web 空間の中でビデオデータなどのリッチデータが急激に増えてきている。これらをいかにアーカイブ形成し、有効利用していくかについての重要性を議論し、平成 21 年度から Web 社会分析プロジェクトを始めた。

現在は図5に示すように、五つのプロジェクトが連携している。この図のように、各々のプロジェクトが独立したものとして位置づけられており、従って、これ自体が連携施策群のひとつのプロトタイプと言える。

以下、各省のプロジェクトの目的、意義などの概略を紹介し、最後に、これらの連携に関する成果について述べる。

##### a. 経済産業省：情報大航海プロジェクト

経済産業省の「情報大航海プロジェクト」では、テキスト、画像、音声、映像のみならず、行動履歴やセンサ情報なども含む多種多様な大量の情報（コンテンツ）の中から、必要な情報を、簡便・的確・かつ安心して收

集・解析・活用し、イノベーションに繋げるための次世代情報検索・解析技術の開発を推進する。

現在、さまざまな分野で多種多様な情報が有効活用されないまま大量に蓄積されている状況のなかで、世界最先端のネットワーク環境を有する我が国において、ユーザのニーズを踏まえた形で、これまで活用されていなかった情報や今後増大していく情報にアクセスし、それらを解析し、活用することが可能になれば、新たなビジネスモデルの創出が可能となる。具体的には、図6に示すように、情報大航海プロジェクトでは、次世代の検索、解析技術の開発、その実証を行うとともに、制度・環境の整備を行っている。

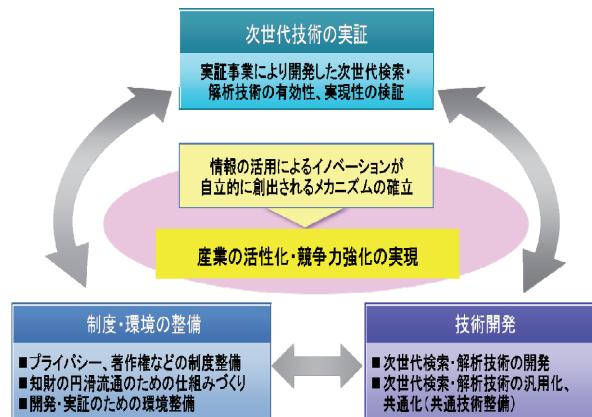


図6 情報大航海プロジェクトの3つの柱

ここで大切なのは、いろいろな分野で言えることであるが、イノベーションの進展には、法制度ならびに環境を整備していくということが重要となる。このプロジェクトを行うにあたっては、この点も踏まえ、積極的に関係省庁に働きかけた結果、著作権法改正の成立に寄与することができたことは、大きな成果といえる。

次に、イノベーションを誘発させる観点から次世代技術の実証を行い、そこで開発した次世代の検索、解析技術の有効性、実現性を検証しながら、きっちりとイノベーションが自律的に行われるようなメカニズムを確立

した。具体的には、図7に示すように、三つのカテゴリ、22の実証事業を行った。

#### A: プライバシーに配慮した未来型サービス

パーソナル情報(プロフィール、行動履歴、趣味嗜好など)を安全に収集、解析することでユーザーの利便性や情報収集力を向上させ、新たなサービスを創出することができる。

#### B: 新しい価値を生む次世代のWebサービス

情報検索等において、ユーザーの利用意図の不明確さや抽象性による問題を解決することまた、リッチコンテンツの利活用を図ることで、市場価値の向上と拡大が期待できる。

#### C: 新たな社会インフラのITサービス

次世代検索・解析技術を活用して、事故や疾患の可能性を事前に察知し、対応する仕組みを構築することによって、様々なリスクを回避・軽減することが可能となる。

図7 主な成果：次世代技術の実証

例えば「プライバシーに配慮した未来型サービス」では、パーソナル情報を安全に収集、解析することでユーザーの利便性や情報収集力を向上させ、新たなサービスを創出すると考えられている。一例として、図8に示すマイ・ライフ・アシストサービスがある。



図8 マイ・ライフ・アシストサービス

このサービスでは、例えば各個人の履歴行動に関するデータを上手く収集・蓄積して解析することにより、新しい気づきや発見に

つながる情報を提供できるものである。このように、さりげなく教えてくれるサービスというの、今後非常に重要になってくると思われる。現在、加入者が300万人を突破したNTTドコモのiコンシェルサービスにおいて位置情報を活用した情報配信が実施されているが、今後このサービスに、マイ・ライフ・アシストサービスによって得られた行動を先読みして有益な情報を提供する技術のノウハウを活用していく予定である。

図9は、このプロジェクトにおけるさらなる成果として、共通技術について示したものである。共通技術には、50を越える技術がある。これらの技術は、先に示した22の実証事業のもとで、いろいろな技術を開発する際に、出来るだけ基盤となる共通技術となるように留意したものである。現在、これらの共通技術は、広く利用していただける形で準備している状況である。

以上のように、情報大航海プロジェクトでは、これらの実証事業を通じて新たなビジネスモデルや共通技術を創出し、その成果を広く一般に共用するオープンな競争環境の整備を促進している。さらに、法制度面・環境の整備に留意しており、この分野における我が国の経済成長の実現に大きく寄与している。

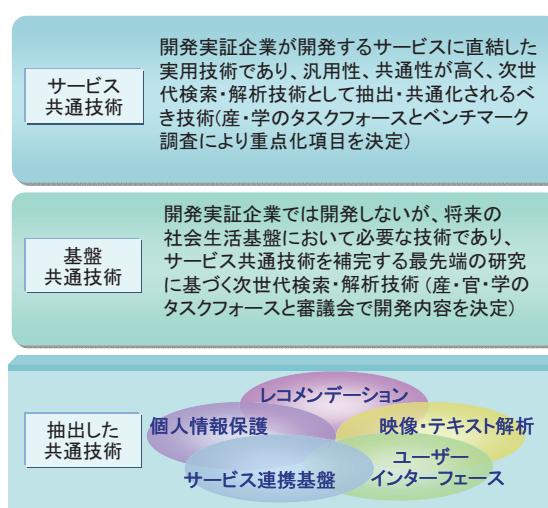


図9 主な成果：共通技術

b. 総務省：電気通信サービスにおける情報信憑性検証技術に関する研究開発プロジェクト

最近では、小学生から大学生に至るまで、宿題で分からぬことがあると、検索エンジンを用いてウェブ情報を調べるなど、あらゆる分野において「不明なことはまずウェブ情報を調べる」ことが当たり前になりつつある。ところが、検索エンジンが誤った情報をトップランクに提示すると、例えば、それがもとで歴史的な重要事実に対する誤った認識をもち、また企業において誤った情報をもとにした製造計画を立て、経済的な損失を被るなど、社会経済的に大きな問題に発展する危険性がある。現在の検索エンジンは、ユーザーの観点から情報を分析して提示していないため、ごみ情報や有害な情報を上位ランクに提示する危険性が多分にある。

特に、ブロードバンド化の進展によって、ネットワーク上の膨大な情報の中には、出所が明らかでないものや有害なもの、正しくないものが混在し、信頼できる情報を得ることが難しくなっている。これらの問題を解決し、誰もが思いのままに、簡単に、しかも信頼して情報を取り扱い、高度に利活用できる環境を実現することが必要である。

そこで、「情報信憑性検証技術プロジェクト」では、ネットワーク上の文字、音声、映像情報について、偽りの情報、信頼性の低い情報などを分析する技術を確立し、より高度な信頼性を伴って情報を提供する環境を実現するための研究開発を推進する。その成果として、例えば、あるコンテンツにアクセスしようとしているユーザーに対して「このコンテンツは60%程度デマである可能性がありますが、それでもよかつたらどうぞアクセスして下さい」というような信頼性情報を提示できるようになれば、ユーザーとしては対象コンテンツの慎重な取り扱いが可能になり、ビジネス上の損失などを減じる可能性が生じる。

情報信憑性検証技術プロジェクトでは、以下に示す二つのテーマを設定し、研究開発を推進している。

#### (1) ウェブコンテンツの分析技術

大量のウェブコンテンツを構成するテキスト、画像、音声、映像情報などの信頼性を、データ解析的な手法に基づき分析する技術を開発する。

#### (2) 意味内容の時系列分析技術

ウェブ上のテキストの集合内に潜在している意見を抽出するとともに、意見間の論理的関係などを抽出する意味理解評価技術や、この意味理解評価技術をもとに各意見の時間的変遷や相互の関係、情報伝播過程などを分析する時系列分析技術を、言語解析技術を用いて研究開発する。

上記において、特に、テーマ(1)の関連で、情報発信者の信頼性を、発信される情報内容やその履歴から評価する技術は未開発である。それに対して、評価用に選定した 10 以上のトピックに関して任意に選択したブログなどの情報発信者の特徴を分析・提示することを実現し、人間による評価と比較して「精度 70%以上」を達成することを目指している。

#### c. 文部科学省：革新的実行原理に基づく超高性能データベース基盤ソフトウェアの開発プロジェクト

前述の二つのプロジェクトに共通することは、今後非常に大規模なデータベースへのアクセスが求められ、従来の実行原理では性能上耐えられない状況になることを想定していることである。

従来、データベース管理システムでは、例えば、

$$\text{残高} = \text{預け額} + \text{利子} - \text{引出額}$$

というような制約を守るために、複数のトランザクションからのデータベースへの同時ア

クセスがあった場合に、それらをある許される順序に慎重に並べて実行することが求められている。

ただし、上記二つのプロジェクトでのデータベースアクセスに関しては、ほとんどがデータの更新を伴わないことなどから、相当レベルの並列度を高めたアクセスが可能であることが見込まれる。そこで、「超高性能データベースプロジェクト」では、大規模データベース管理システムの超高性能化技術の創出を目指し、非順序実行原理に基づく超高性能データベースの基盤となるソフトウェアの開発を推進している。特に、図 10 に示すように、今後 3 年後に 10 倍、5 年後には 100 倍程度の性能改善を目指している。

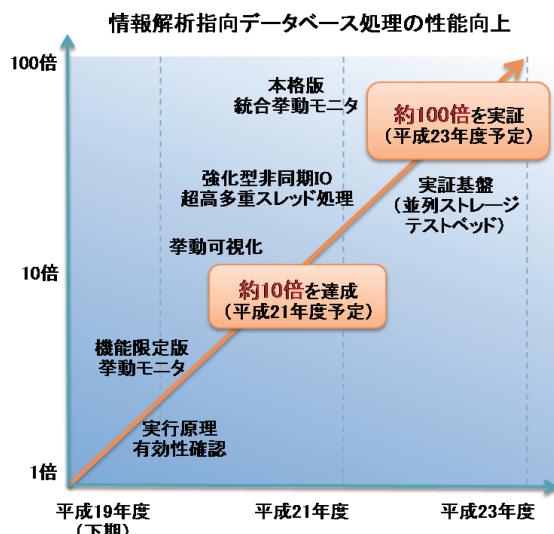


図 10 成果目標

これまでの主な結果として、データベースの分野においてよく使われているオープンソースである DBMS (MySQL)、商用のデータベース (HiRDB) において実験を行い、現時点で、その性能が約 20 倍の性能向上を達成しており、平成 21 年度の目標である約 10 倍を大きく上回る成果を既に達成している。

d. 文部科学省：多メディア Web 解析基盤の構築及び社会分析ソフトウェアの開発プロジェクト

平成 21 年度から新しい施策として「多メディア Web 解析基盤の構築及び社会分析ソフトウェアの開発」を開始した。この施策は、いわゆる社会学、言語学、リスク管理、マーケティング等の多様な社会分析ニーズに応じるために、膨大な多メディア Web 情報を収集し、蓄積して、多様な解析を可能にするということを目指している。最終的に社会分析のソフトウェアの研究開発、ならびに実証を行うものである。

現在、大きな流れとして、動画トラフィックの増加、すなわち、ビデオデータが相当増えてきており（ここ数年間で 6 倍）、これをどのように使える形にするのか、また、実世界情報とサイバー空間のデータは、相互に影響を及ぼし合うということが広まってきている。このような中で、多メディアに関する Web の基盤技術、それから解析要素技術を上手く連携させることによって、実際に社会分析ソフトウェアを開発する。

特にこの研究開発では、多メディア Web アーカイブ中で観測可能なイベントや話題の発生時点であるとか、それらの発展、反響等の時系列追跡を正確に行うこととし、これらをリーズニング解析と呼び、社会分析の軸とすることを考えている。

現在、ある程度までは使える段階になっており、今後、実証実験によりその精度を上げていく。

e. 補完的課題：センサ情報の社会利用のためのコンテンツ化プロジェクト

現在、国内では平成 18 年度から開始された国家事業である u-Japan 構想によって、「ユビキタスネット社会」を平成 22 年までに実現することを目標に据え、体系的な情報通信技術政策を推進中である。そのような状況のなかで、近年、センサネットワーク技術が急速な勢いで発展しており、それによってもたらされるさまざまな応用技術に対して大きな期待が寄せられている。特に、日常の生活空間からセンサで得られた大量の情報をウェブ化し、知的情報アクセス基盤技術を活用する技術は、高度に進化した「ユビキタスネット社会」を構築していく上で重要である。

連携施策群で推進される情報大航海プロジェクトにおいては、ユーザ主導型でありビジネスに近いところの技術開発が主となっている。屋内外のユーザがどのようなコンテキスト（状況依存）で情報を要求しているかを知るためにには、公共環境に設置されたセンサなどからも情報収集する必要があり、まだまだ基礎的な研究が必要な部分である。つまり、連携施策群の既存の研究プロジェクトによつて、日常の生活空間なども含めたさまざまな情報空間から獲得されるセンサ情報などを扱うことは全く不可能ではないが、信憑性検証技術をはじめ、センサ情報の利用技術については基礎的な研究も含め不十分である。

そこで、本連携施策群の成果を有効に活かすためには、各プロジェクト終了後の技術マップにおいて、平成 24 年頃を想定した次世代情報環境の技術開発を行うことが非常に重要なってくる。

図 11 に示すように、センシングウェブ技術に関する研究を補完的課題として推進し、拡大する領域の技術開発を有機的に連携しつつ、先行して行うことで、各省施策の技術開発を加速することが可能となる。具体的には、

「センサ情報の社会利用のためのコンテンツ化」プロジェクト（代表機関：京都大学）を推進しており、特に、本課題を効果的に実証するために公共環境での実証実験場所（京都：新風館）を確保して研究推進が行われている。

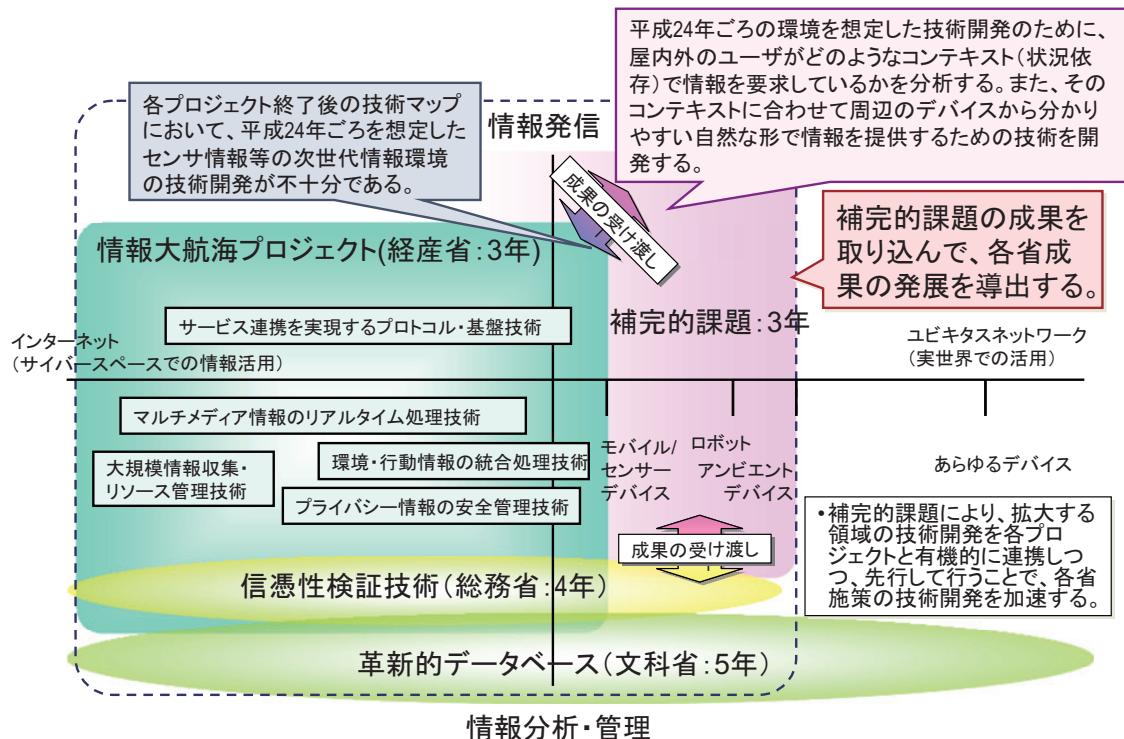


図 11 補完的課題の位置づけ

本プロジェクトでは、「情報共有」、「アクセス管理」、「情報活用」を三つの柱として、以下のようなテーマを設定している。

- (1) センサ情報を共有するメカニズム（情報共有）
- (2) センサ情報特有の問題への対処（アクセス管理）
- (3) 分散したセンサ情報の検索・提示法（情報活用）

このプロジェクトの成果は、その研究開発途中から各省施策にフィードバックすることも可能であり、研究開発基盤となることも予想され、直ちに各プロジェクトに活用できる。したがって、補完的課題として行うことで相乗効果を高めることができ、大きな期待が寄せられている。これまでに、図 12 に示すように、実環境として京都の新風館（ショッピングモール）で実証実験を行ってきており、利用者からの受容性評価で 8 割強という高い評価を得ていることからも十分期待できるものである。

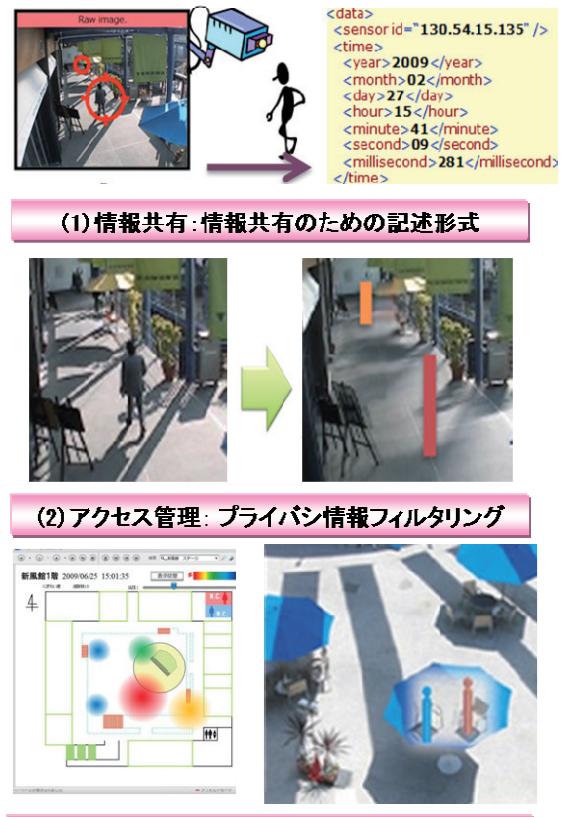


図 12 新風館での実証実験

f. 「情報の巨大集積化と利活用技術開発」施策の技術要素間連携

これまでに述べた各施策に関して、どのように連携してきたかについて述べる。本連携群における各施策を進めるにあたり、先に述べた運営体制のもと、タスクフォースを定期的に開催し、常に連携を意識した開発となるように留意してきた。このタスクフォースにおける議論を踏まえ、図13に示す各施策間の技術要素間連携図、いわゆる連携マップを作成した。各施策は、この連携マップを基に、施策間における具体的な連携シナリオを想定しながら、各施策の開発を進めるようにした。各施策間で密に情報交換を行いながら、かつ施策の独立性も保ちつつ効率的な連携が図れるように進めてきた。

図13に示すように、例えば、情報大航海プロジェクトとセンシングWeb間では、実証実験の成果に関して、パーソナル情報の利活用という観点より密に連携し、その結果は、後述の「4.3.1」項に示す活用例「パーソナル情報保護・解析基盤」に活かされている。また、同じく情報大航海プロジェクトと情報信憑性検証技術においても、お互いに情報交換しながら検討を進め、その結果は、「4.3.5」項に示す活用例「プログラミングサービス」に活かされている。さらに、他の連携群との連携に関しても情報交換を進め、平成20年度に終了した次世代ロボット群と補完的課題において、特にセンサとのインターフェース仕様に関して共通化を図り、その結果に関しては、「4.3.6」項の「センシングWeb」に組み込まれている。

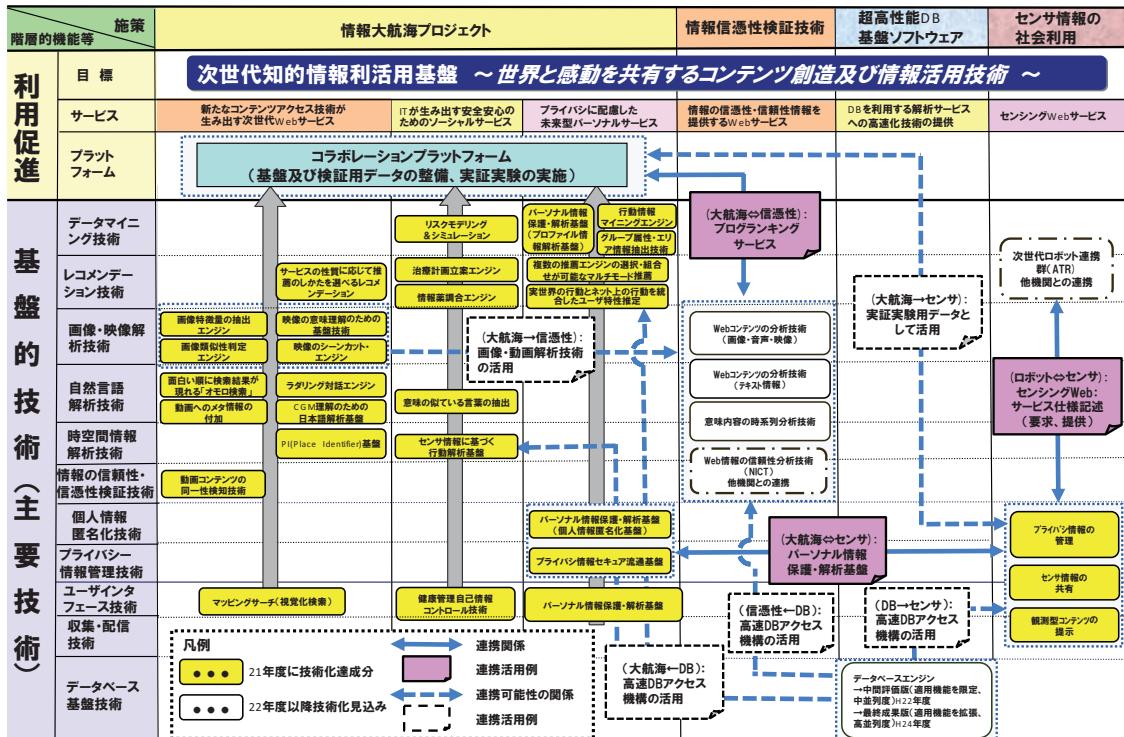


図13 「情報の巨大集積化と利活用技術開発」施策の技術要素間連携図

## 5. 連携群の活動における3年間の総括

図14、15は、この3年間の活動の総括をとりまとめた結果である。これらの結果から、当初の目的以上のいろいろな成果、例えば、超高性能データベース基盤に代表されるように、当初目標値を大幅に上回る成果を実現した。また、施策のさらなる発展として、新規施策の追加、最先端研究開発支援プログラムに採択など適用領域の拡大化を行ってきた。また、法制度面においては、著作権法改正の成立に寄与した点などが挙げられる。

国際的な活動に関しても、欧州連合(EU)を中心とし、視察を行うなど、国際連携を模索してきた結果、例えば、経済協力開発機構(OECD)代表部より、個人情報、プライバシー情報の利活用に関する議論を日本がリードできる可能性があるとの評価を得たことは、非常に大きな成果である。

以上により、連携群の活動の総括として、当初計画以上の成果を十分に達成した。

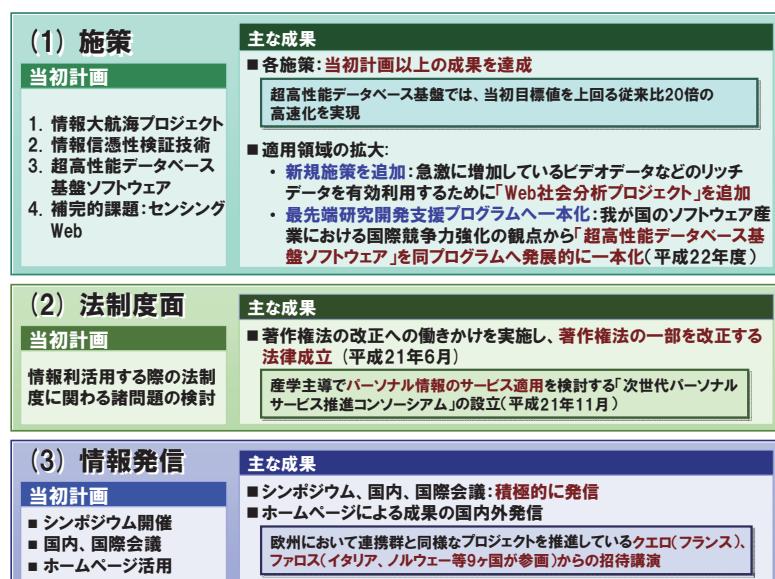


図14 主な成果(3年間の総括)

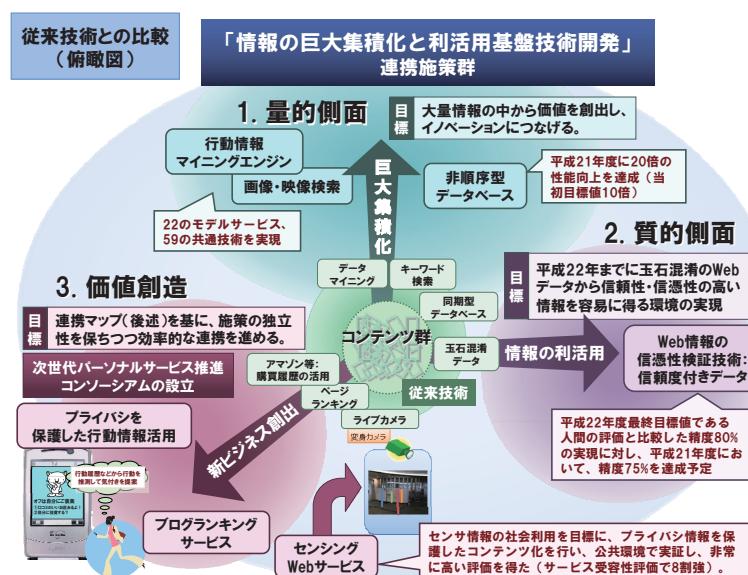


図15 主な成果(従来技術との比較)

## 6. 今後の展開

今後の展開として、まだまだ取り組むべき課題は多い。平成 21 年度から始めたプロジェクトを例にとっても、ビデオデータなどのリッチコンテンツの下で知的情報基盤のさらなる推進が必要となると考えられる。本連携施策群のプロジェクトは、開始当初より 3 年間としており、5. の総括で述べたように、そのベースはしっかりと築くことが出来たことから、進行中の施策も含め、さらなる進展が期待できる。また、法制度についても、今後、第 3 期以降の科学技術基本計画策定の際に検討すると思料される科学技術・イノベーションというキーワードは、従来の技術開発に加え、新たに社会科学との連携、法制度との連携が必要となる点に関しても、このプロジェクトがその先駆け的な意味を果たせたと言える。

本連携群の今後に関しては、今年度で、情報大航海プロジェクトとセンシング Web が終了する。他に、継続するプロジェクトが三つあるが、これらに関して、今後、何らかの形で議論の場を設け、これら五つのプロジェクトに関して、有機的な連携を継続して取ることが重要である。このようなプロジェクトを継続する機会を設けることにより、国際競争力のさらなる向上に寄与しうることになり、今後の大きな課題である。

以上のように、本連携施策群は、イノベーション創出と法制度という観点からも大きな意義を有している。情報大航海プロジェクトにおいて、多種多様な大量の情報（コンテンツ）を解析してデータベース化し、ビジネス展開などに再利用するために、元となるコンテンツに関する版権、著作権などの問題に取り組み、前述のように著作権法改正に少なからず寄与することができた。また、公共環境からのセンサ情報を対象にする場合は、顔映像をはじめとする個人の肖像権に抵触しないように研究開発が求められる。このようにイ

ノベーションを起こすために、法制度による拘束が障害になる場合がしばしばあり、それらの緩和を働きかける活動なども、一方ではますます重要になってくるであろう。

### **3. 連携施策群の活動報告**

- 3.1 情報大航海プロジェクト：経済産業省
- 3.2 電気通信サービスにおける情報信憑性検証技術に関する研究開発：総務省
- 3.3 革新的実行原理に基づく超高性能データベース基盤ソフトウェアの開発：文部科学省
- 3.4 多メディア Web 解析基盤の構築及び社会分析ソフトウェアの開発：文部科学省
- 3.5 センサ情報の社会利用のためのコンテンツ化  
(補完的課題)：京都大学(研究代表機関)

### 3.1 情報大航海プロジェクト

経済産業省商務情報政策局 情報処理振興課長  
東條 吉朗

#### 概要

情報大航海プロジェクトは、最終年度である3年目、正にプロジェクトの終了にさしかかっている。本稿では、情報大航海プロジェクトの目的、概要、成果、今後の展開等について述べる。

#### 1. 背景

近年、Web上のデジタル情報にとどまらず、企業内で生成されるデジタル情報、GPSを活用した位置情報・移動情報、個人の消費行動に関する情報等、社会活動のあらゆる場面で大量の情報が創出・蓄積されており、世界中で情報量の「爆発」が起こっている。情報大航海プロジェクトはこの情報爆発の状況の中で、これらの情報をいかにして有効活用し、新たなビジネスやイノベーションの創出、安全安心で豊かな国民生活につなげていくかという問題意識の下にスタートした。

図1に示すとおり、社会活動において創出・蓄積される情報は、平成20年(2008年)には約500エクサバイト(EB:1EB=10億GB)にのぼり、これはさらに指指数関数的に増大していると言われている。このような状況にお

いて、Google等の直接検索で活用されているWebページの情報は0.3EBであり、Web上には存在しているが検索の対象となっていないDeepWebの情報量を合わせても15EB程度である。その一方で、センサ情報であるとかRFIDのタグ情報であるとか、そういうところから読み取られた情報等、ほとんどの情報は非Webのところにある。これらは蓄積されている全情報量の95%にのぼり、整理・構造化されないまま、すなわち、上手く活用出来る状態になっていないまま、サイバー空間に存在しているのである(図2)。

このサイバー空間上で手つかずのまま置いておかれた非Webの大量の情報も含め、サイバー空間上の大量の情報の中から価値を創出し、イノベーションにつなげようというのが情報大航海プロジェクトの狙いである。

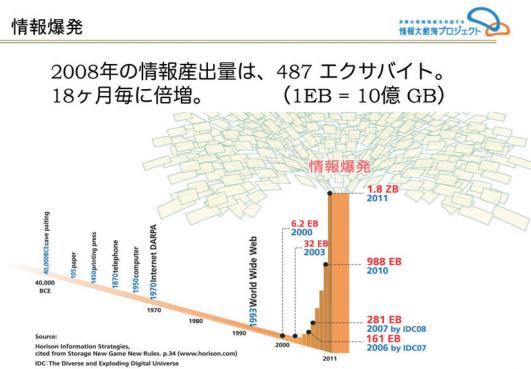


図1 情報爆発

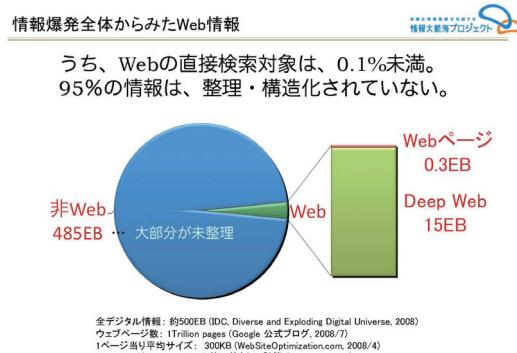


図2 情報爆発全体からみたWeb情報

## 2. 取組み

我々はこのプロジェクトの3年間、以下の取組みを一体的に実施してきた（図3）。

- Web 上の情報にとどまらず多種多様な大量の情報の中から必要な情報を的確に検索・解析する次世代検索・解析技術を開発し、モデルサービスにおいて実ビジネスのテストベッドを構築して、新規ビジネスを先導する。
- モデルサービスにおいて開発された技術を共通・汎用化するとともに基盤として必要な技術を新規開発することにより共通の技術基盤を構築・運用し、これらの技術の普及を図る。
- モデルサービスを通じて制度的課題を洗い出し、技術の市場展開に必要な環境整備を行う。

さらに3年目の平成21年度は、開発した技術の国際標準化等、成果の国際展開にも注力し、開発技術の普及や新規市場の創出を目指してきた。

まずa.について、3年間で情報大航海プロジ

エクトにおいて行ったモデルサービスは表1のとおりである。延べ22のモデルサービスを実施し、これにより新規サービス創出に向けたノウハウを得ることができた。

また、b.については、これらのモデルサービスやそれ以外の新規サービスにおいてオープンに利用できる共通の技術基盤として、モデルサービスの実施のために開発された「サービス共通技術」とサービス共通技術を補完する「基盤共通技術」の合計59技術を開発した（図4）。これらの共通技術や実証用データ、映像コンテンツ等を整備し、共有・検証する環境として、コラボレーションプラットフォームの運用と改良を行ってきた。このコラボレーションプラットフォームは、高性能の設備を要する開発・検証作業を行う必要があるもののそのような環境を持たないプロジェクト関連企業において必要な環境として活用されるなどした。さらに、開発した共通技術については、「PI（Place Identifier）基盤」が国際標準化機構（ISO）の新作業項目提案（New Work Item Proposal）として承認されたほか、「パーソナル情報保護・解析基盤」案（New Work Item Proposal）として

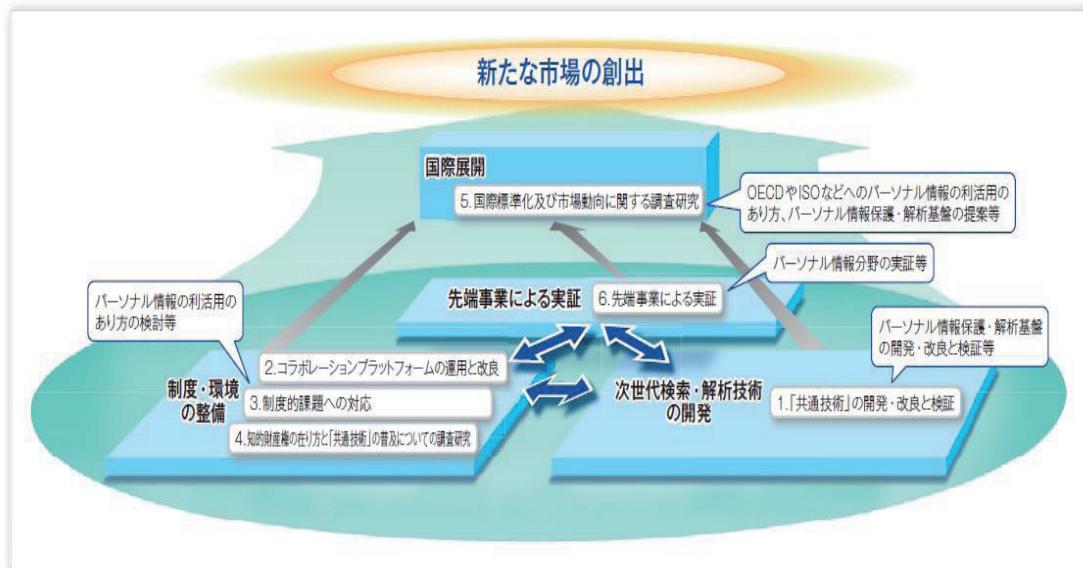


図3 情報大航海プロジェクトにおける三位一体の取り組み

承認されたほか、「パーソナル情報保護・解析基盤」のISO提案に向けて取り組んできた。このほか、クエロ(Quaero)、コーラス(Chorus)等、欧州の関連する研究開発プロジェクトと積極的に情報交換・連携を行ってきた。

c.については、著作権制度やパーソナル情報の利活用のあり方について制度的課題への対応を行ってきた。著作権制度については、

法改正に向けた継続的な働きかけを行った結果、平成21年6月12日に「著作権法の一部を改正する法律」が成立し、検索サービス事業の実施が可能となった。パーソナル情報の利活用のあり方については、有識者による研究会を設けて検討を行い、検討の方向性等について経済協力開発機構(OECD)と情報交換・連携を行うなどした。

表1 情報大航海プロジェクトにおいて実施したモデルサービス

実証事業	開発実証企業
平成19年度 (10実証事業)	
ラダリング型検索サービス	沖電気工業株式会社
サグールテレビ	チームラボ株式会社
View サーチ北海道	株式会社データクラフト
時空間情報マイニングサービス	株式会社NTTデータ
すこやかライフサポートサービス	財団法人国際医学情報センター
新総合安全運航支援システム	株式会社日本航空インターナショナル
メガリサーチ	モバイルジャッジ株式会社
マイ・ライフ・アシストサービス	株式会社NTTドコモ
交通系非接触ICカードを活用した連携型サービス	東京急行電鉄株式会社
プロファイルパスポート事業	株式会社ブログウォッチャー
平成20年度 (8実証事業)	
マイ・ライフ・アシストサービス	株式会社NTTドコモ
ラダリング型検索サービス「ラダサーチ」	沖電気工業株式会社
地域活性化を支えるe空間サービス～ぷらっとPlat～	株式会社エス・ピー・シー
「ここなら」コミュニケーションサービス	株式会社ジー・サーチ
多言語対応動画アプリケーションプラットフォーム	株式会社角川メディアマネジメント
異業種連携IDによる流通サービス	メディアラグ株式会社
次世代解析技術を活用した携帯情報端末による健康管理	株式会社キューデンインフォコム
新総合安全運航支援システム	株式会社日本航空インターナショナル
平成21年度 (4実証事業)	
マイ・ライフ・アシストサービス	株式会社NTTドコモ
ラダリング型検索サービス「ラダサーチ」	沖電気工業株式会社
次世代解析技術を活用した携帯情報端末などによる情報循環方式の健康管理	株式会社キューデンインフォコム
新総合安全運航支援システム	株式会社日本航空インターナショナル

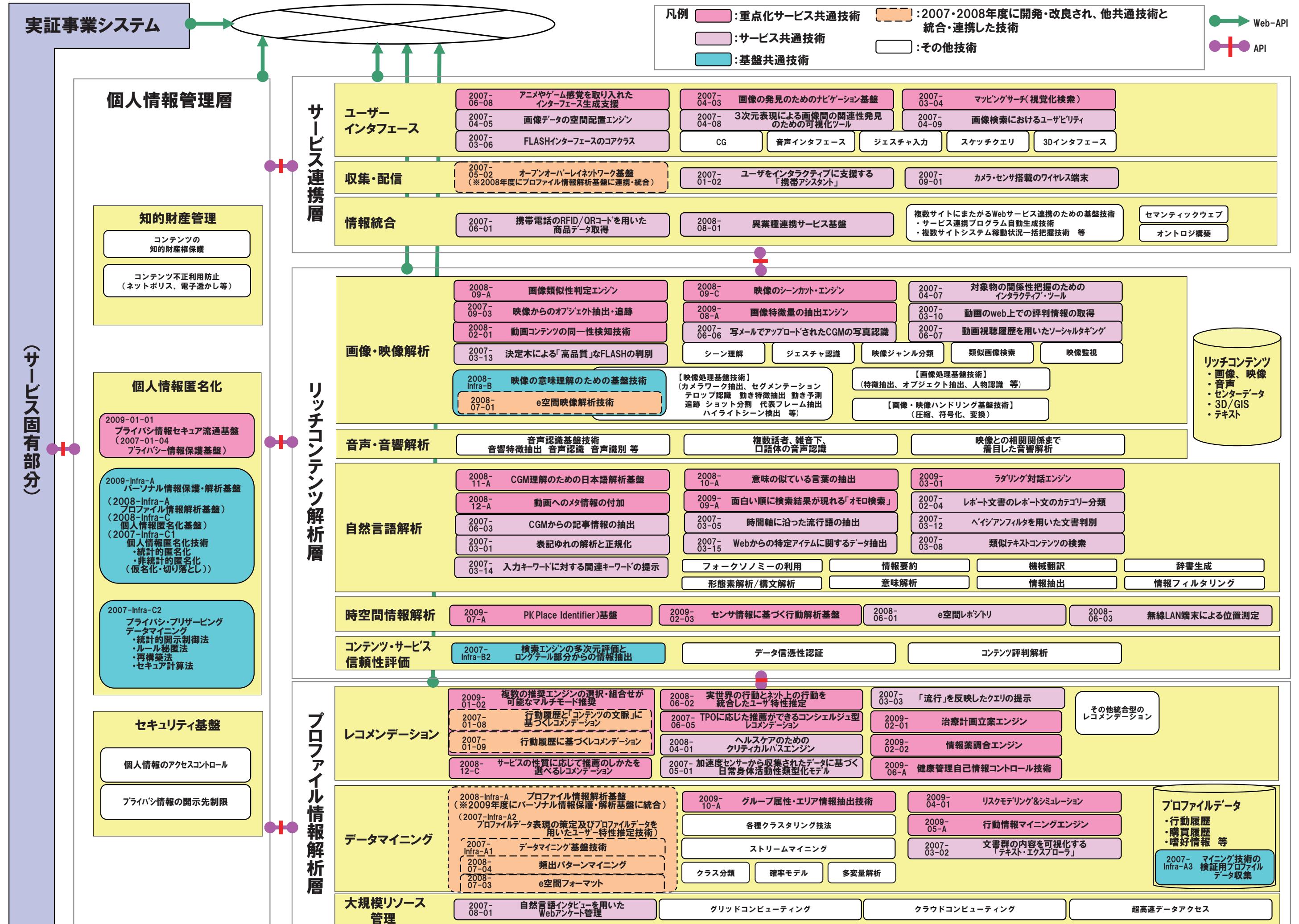


図4 基本技術アーキテクチャ（技術マップ）

### 3. 成果

情報大航海プロジェクトは前述のとおり社会のありとあらゆる情報を利活用しようという取組みであることから、対象とする分野が多岐にわたる。

ここでは、対象として以下の分野を抽出し、分野ごとにその成果を述べる。

- a. 行動履歴に基づくパーソナルサービス
  - b. センサ情報を活用した健康・疾病管理
  - c. ヒヤリハット情報解析を通じた安全性・信頼性向上
  - d. 画像情報解析によるコンテンツ市場活性化
    - (a) 行動履歴に基づくパーソナルサービス
- パーソナルサービスの分野では、例えば、NTT ドコモが平成 20 年度から i コンシェルサービスを始めた。i コンシェルサービスにおいては、GPS による位置情報を活用した広告配信等を行っている。現在のサービスにおいては、立ち寄った場所に応じて広告を配信するという単純な機能が実装されているところであるが、情報大航海プロジェクトにおいては、位置情報をずっと繋げていって行動履歴を収集し、行動パターンを解析することによって、もっと精度の高いレコメンデーションをユーザの行動を先読みして提供できるようなプラットフォームを作ろうとしてきた。実際に i コンシェルサービスそのものは、平成 20 年 11 月から開始され、平成 21 年度末時点では、加入者が 300 万人を超え、100 億円程度の市場になるとみられている。さらに将来的には 1000 万人加入、300 億円程度の売上げを目指しているが、このサービスに情報大航海プロジェクトの成果を活用し、次世代のパーソナルサービスに繋げていこうとしている。この行動履歴解析によるレコメンドサービスの事業化が、情報大航海プロジェクトへの参加によって 2~3 年早くなったとのことである。

また、このサービスは、プラットフォームを提供する NTT ドコモのような事業者だけではなく、プラットフォーム上で新しいサービスを提供する企業にも広く開かれたプラットフォームになるはずである。NTT ドコモの事業に関して言えば、NTT ドコモにおける当該事業そのものの売上が 300 億円だとすると、そのプラットフォーム上に更に数倍規模の新しいサービス市場が開かれる展望が見えてきたところである。

このほか、平成 20 年度に情報大航海プロジェクトで実施されたモデルサービスを元に、平成 21 年度から、同じ経済産業省のプロジェクトとして、デジタルサイネージ等、IT の活用によって地域の商圏を活性化しようとする試みが情報大航海プロジェクトからスピンオフして実施されている。

この分野の技術面に関しては、こういったサービスを支える技術基盤として、様々な情報を収集し、そこからパターンを抽出する技術「プロファイル情報解析基盤」や行動履歴等を集合匿名化し、プライバシ配慮する形で情報基盤として共有・利活用するための技術「個人情報匿名化基盤」を開発し、オープンな技術として普及を図っている。

技術基盤の裏側で、制度環境の整備も必要である。ご存じのように個人情報の保護については法制度が既に確立しているものの、パーソナルサービスを実施していく中で、どう個人情報を取り扱い、匿名化していくかに利活用するかについては、ある程度手続きが明らかでないとプライバシ保護の観点からビジネスリスクが存在するものと考えている。実際、事業者からは、容易にそのようなデータの利活用に手を出しにくいとの声も聞こえている。情報大航海プロジェクトにおいては、技術基盤のみならず、個人情報、匿名化情報をどのような手続きで、どのようなデータ形式にすれば、プライバシ保護に抵触しないか

ということについて検討を行っており、その検討結果については、各省の個人情報保護ガイドラインとの摺り合わせを行ったり、あるいはパーソナルサービスを実施しようとする業界団体の自主的なガイドラインという形で結実させたりすることを考えている。また、各種サービスやデータは容易に国境を越えて広がっていくことから、我々の検討内容については、前述のようにOECD等を通じて国際的な連携を図り、国際的に見ても合意の得られるプライバシ配慮を行った健全な市場を作りたいと考えている。

さらに、パーソナルサービスの分野においては、数十の企業、研究者、法律家等の方々により、「次世代パーソナルサービス推進コンソーシアム」が設立されたところであり、我々としては、こうしたコンソーシアムを含む民間のイニシアチブが、匿名化情報の利活用のためのガイドラインの主体や情報大航海プロジェクト終了後の基盤技術の普及の主軸となり、パーソナルサービスの新たな市場創出を牽引してくれることを期待している。

## (b) センサ情報を活用した健康・疾病管理

健康・疾病管理の分野では、ユーザが持った加速度センサ等で収集した情報を解析することにより、その人がどのような姿勢でどのような行動をしているのかを判別し、医療機関や健康サービス機関が生活習慣病の治療や健康指導としてリアルタイムにユーザに適切な情報を提供することにより、生活習慣病の機動的かつ効果的な治療や予防を実施する「情報薬」のサービスを行っている。

ご存じのように、生活習慣病は先進諸国で問題となっており、各国の医療費を圧迫している現状にある。我が国における糖尿病患者は、予備軍を含めると2200万人に上るといわれており、平成27年度までにメタボリック・シンдромの該当者および予備軍を平成20年度比で25%減少させ、医療費を2兆円抑制するため、平成20年度からいわゆる「メタボ健診制度」が開始されたところである。このように、こうした健康・疾病管理サービスというのは、社会保障負担の緩和に寄与するも



図5 パーソナルサービス分野の成果と今後の展開

のであるほか、地域の医療機関の許容量といった制約をある程度緩和し、その地域の住人の便益を高める期待も高い。情報大航海プロジェクトのモデルサービスは、平成20年度は九州において生活習慣病予防サービスとして実施したところだが、平成21年度は福岡市医師会と協力し、糖尿病治療としてサービスを実施している。サービスを受ける患者は、「電車を一駅手前で降りて歩きませんか。」といった行動解析結果に基づく生活改善情報を受け取るほか、「最近手足のしびれを感じることはありましたか。」など、合併症の早期発見につながる情報も受け取る。また、かかりつけ医の側にも、治療計画立案エンジンを用い、糖尿病治療のガイドラインに沿った治療ができるようサポートを行っている。このように、適切なタイミングで情報提供するサービスを開拓することにより、「情報薬」の検証を行おうとしているのである。

このサービスから生まれた共通技術である

「センサ情報による行動解析基盤」は、「走る」、「座る」、「電車」、「喫煙」等の22の行動パターンを89%の精度で判定するという成果を得ているが、この技術は国際的にも関心を集めている。例えば、先般、カナダのレッド・エンジン・ヘルス社と情報大航海プロジェクト参加企業の業務提携が決定され、生活習慣病が社会問題化しているカナダのアルバータ州の対策プログラムに、この技術を活かしていくこととなった。

また、国内においても、東京や静岡の医師会がITを利活用した健康管理について検討を開始するなどしており、この分野における我々と同種の取組みが国内外を問わず広がりつつある。

もちろん、医療や健康の分野においては、健康保険の適用等、制度面でクリアすべき課題が多くあるが、有望な市場につながる技術、そしてビジネスモデルが提示できつつあると言えるだろう。

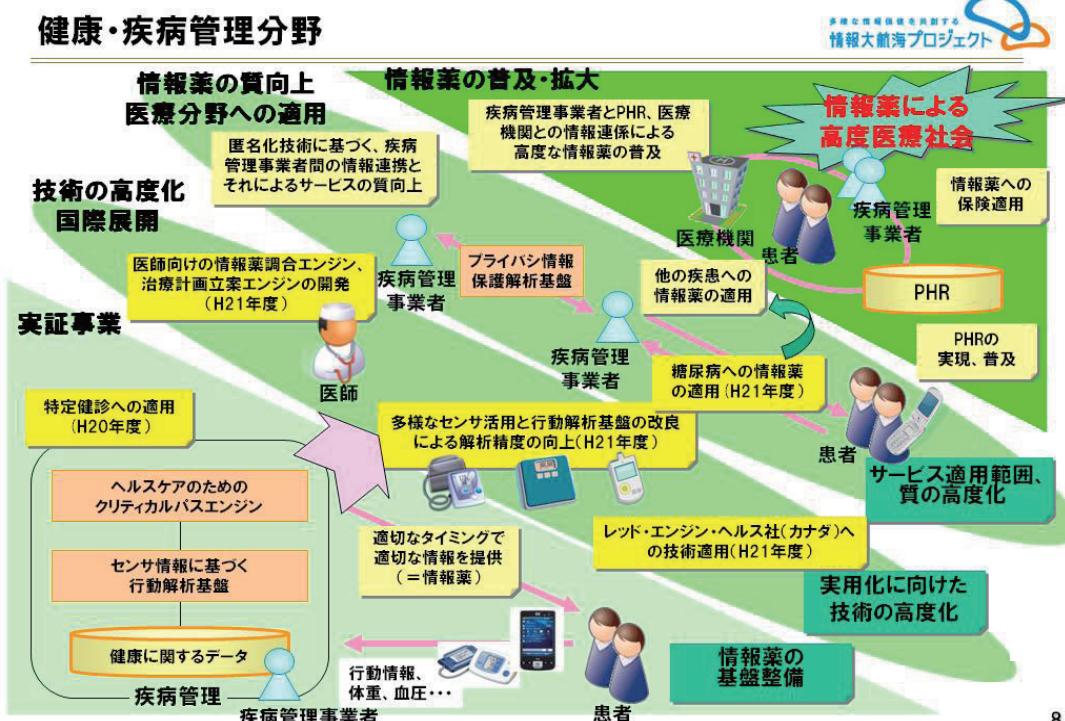


図6 健康・疾病管理分野の成果と今後の展開

## 安全安心分野

情報大航海プロジェクト



図7 安全安心分野の成果と今後の展開

### (c)ヒヤリハット情報解析を通じた安全性・信頼性向上

安全安心分野では、JALを中心としたチームで、実際のヒヤリハット情報のレポートをデータマイニングすることにより、ヒヤリハット事案の因果関係、事案当時の諸状況を自動的に解析し、安全対策立案を支援するという取組みを行っている。技術的には、自然言語解析技術等を開発・活用し、解析に要する期間について、人間の手で5ヶ月必要な作業を同程度の精度で2ヶ月に短縮することが可能となり、安全対策の立案をこれまでの倍のスピードで行うことができるようになってきている。

今年度はさらに、レポートのテキスト情報に加えて事案発生時における気象等様々な条件に関するセンサ情報を分析の対象としたところであり、センサ情報をリアルタイムに分析し、安全対策にその場でフィードバック

していくことが将来的に可能かということについて展開をしつつある。

このモデルサービスについては、航空業界において展開可能なのはもちろんのこと、同様の取組みを他の社会インフラの分野に展開することが期待される。実際、本年の取組みにおいて、エール・フランス等他の航空会社への展開に向け、多言語化や外国語のレポートでの実証も行っているところであるほか、国内の鉄道会社や電力、ガス会社への展開に向けて、コンソーシアム形式により各社と連携しつつ実証を進めているところである。

### (d)画像情報解析によるコンテンツ市場活性化

現在の一般的な検索サービスにおける画像検索や動画検索においては、その画像等に紐付けされたタグ情報を検索して結果を提供しているところであるが、情報大航海プロジェ

クトにおいては、画像そのものを解析して実際の画像の類似性等から検索評価を行う技術や大量の動画コンテンツを対象として既存のコンテンツとの同一性を検知する技術等の開発・実証を行った。これらの技術は、画像・動画コンテンツへのユーザのアクセスを容易にするほか、創作活動に必要な材料を収集するなどにより創作活動を支援する。また、動画検知技術により、コンテンツ市場における権利保全が進むものと考えている。

この分野において、モデルサービスの実施という観点からは、情報大航海プロジェクトの中では、平成 20 年度までほぼ終了しているところであるが、そのために開発された共通技術については普及が進み、現在、商用等における活用事例が 100 件を超えており、6 兆円産業であるデジタルメディアの中で、数割のシェアを占めるテレビ業界、映画業界等の企業においても、情報大航海プロジェクトにおいて開発された共通技術を活用し、彼らのビジネスに実装している。

このほか、当省のプロジェクトにおいて、コンテンツホルダーが連携して、中国の違法動画サイト等の違法コンテンツを検索し、適切な対処を行うという実証実験を開始したところであるが、この実証実験の対象技術として、情報大航海プロジェクトの開発した共通技術が採用されている。

この分野の制度に関しては、前述のとおり、平成 21 年 6 月に著作権法の改正が行われ、検索サービスや情報解析のための著作物の複製について、ある程度柔軟な取扱いができるようになった。これも情報大航海プロジェクトの中で、産学連携を深めながら、平成 19 年から働きかけた結果がこの改正に結びついたものと考えている。著作権制度については、今般の改正著作権法の下での実際のビジネスの展開がスムーズになるよう環境整備を行うとともに、著作権法改正後にも存在する問題点について整理の上、引き続き働きかけを続けていきたい。

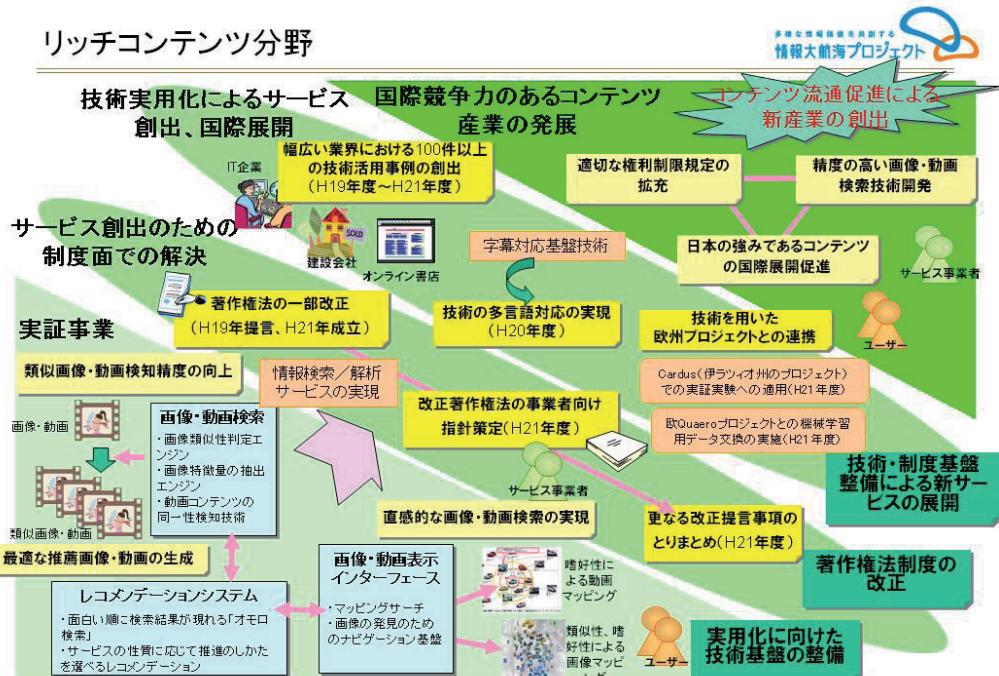


図 8 リッチコンテンツ分野の成果と今後の展開

#### 4.まとめ

情報大航海プロジェクトは、既存のサービスで活用されてこなかった非Webを含む大量の情報を有効活用することにより、新たなビジネスやイノベーションを創出し安全安心で豊かな国民生活につなげるためのプロジェクトであり、この目的を達成するため、「新ビジネスの先導」、「基盤技術の開発・普及」及び「制度的課題の抽出と環境整備」の三つの取組みを3年間かけて一体的に実施してきた。

その結果、新ビジネスの先導の観点から言えば、パーソナルサービス分野や健康・疾病管理分野では実際の事業化につながる形で展望がひらけてきたほか、社会インフラの安全性・信頼性向上やコンテンツ産業の活性化についても一定の貢献をなすような基盤ができるところである。また、基盤技術の開発・普及については、開発した共通技術について100件以上の商業利用につながっているほか、プロジェクト終了後にも次世代パーソナルサービス推進コンソーシアムやヒヤリハット情報のテキストマイニング普及のためのコンソーシアム等を通じて、こうした共通技術の普及が行われるという展望がひらけている。さらに、制度的課題については、検索サービス等に関する著作権法の改正が情報大航海プロジェクトをきっかけに行われたほか、今後もパーソナルサービス推進のための情報の利活用のあり方の検討等、プライバシとデータマイニングの両立についても引き続き取り組んでいく。パーソナルサービスのための情報利活用については、匿名化エンジン等の国際標準化と制度面でのOECD等との国際協調を推進し、ビジネスモデルが日本にとどまらず世界の市場に飛躍できるようにしていきたいと考えている。

最後に、情報大航海プロジェクトは3年目の終了を間近に控え、上に述べた形で、データマイニングによるイノベーションの可能性について一定の役割を果たしつつあるが、情

報爆発がさらに継続していく中には、この情報の有効活用の必要性が今後ますます高まっていく一方で、情報処理について高速化や低炭素社会化のための効率化といった点が課題になってくると考えられる。我々としては産学と連携しながら引き続き新規ビジネス創出の後押しをしつつ、この「情報の大集積化と利活用基盤技術開発連携群」をきっかけに深化することができた文科省、総務省等の関係省庁等との連携を活かしてデータマイニング利活用の基盤となる情報処理の技術基盤についても協力していきたい。

### 3.2 電気通信サービスにおける 情報信憑性検証技術に関する研究開発

総務省総合通信基盤局電気通信事業部電気通信技術システム課企画官  
安全・信頼性対策室長 川村 一郎

#### 概要

ネットワーク上の文字、音声、映像情報について、偽りの情報、信頼性の低い情報等を分析する技術を確立し、信頼できる情報を提供することで、誰でもが思いのまま、簡単に信頼して、コンテンツを取り扱い、高度に利活用できる環境を実現する。

#### 1. 背景・目的

ブロードバンド化の進展によって、ネットワーク上の膨大な情報は、出所が明らかでないものや有害なもの、正しくないものが混在し、信頼できる情報を得ることが難しくなっている。

このため、これらの問題を解決し、誰でもが思いのまま、簡単に、信頼してコンテンツを取り扱い、高度に利活用できる環境を実現することが必要となっている。

また、本研究開発は、世界最先端のIT国にふさわしい技術水準を確保し、高度情報通信ネットワークの健全な利活用を推進するといった点からも必要である。

#### 2. 研究内容

情報信憑性検証技術を確立するため、平成19年度から平成22年度までの4年計画で以下の技術の研究開発を実施している（図1参照）。

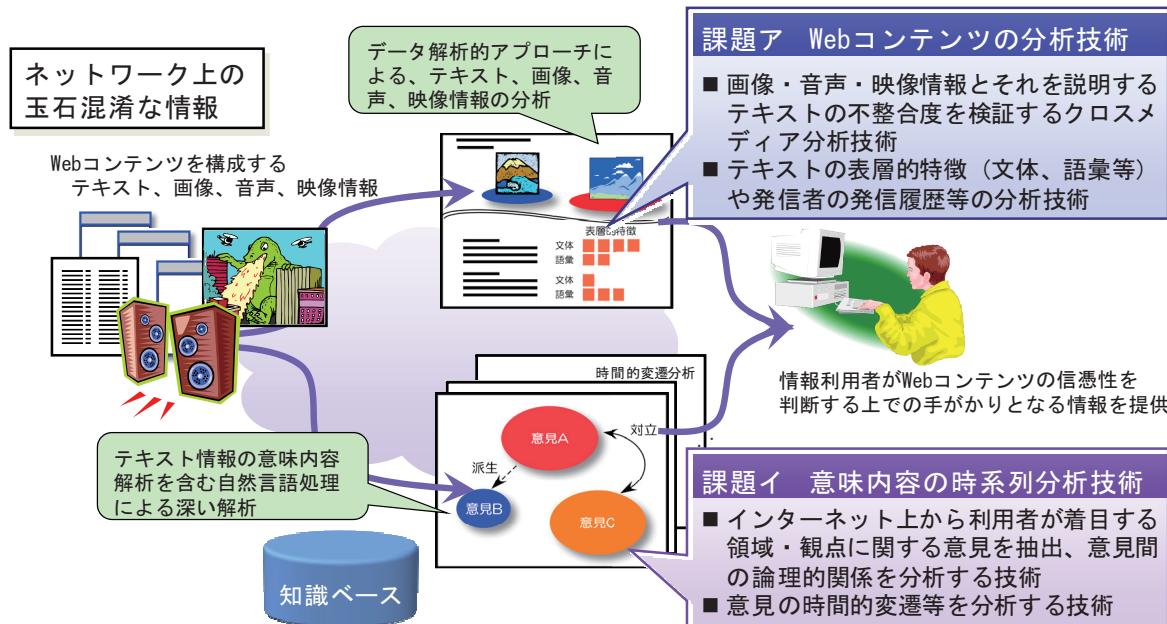


図1 情報信憑性検証技術における研究開発内容

- a. Web コンテンツの分析技術（課題ア）  
研究者：京都大学大学院（研究代表）、兵庫県立大学、京都産業大学、（株）きざしカンパニー、（研究協力：ヤフー株式会社）
- (1) 画像・音声・映像情報とそれを説明するテキストの不整合度を検証するクロスメディア分析技術
- (2) テキストの表層的特徴（文体、語彙等）や発信者の発信履歴等の分析技術
- b. 意味内容の時系列分析技術（課題イ）  
研究者：日本電気株式会社（研究代表）、奈良先端科学技術大学院大学、横浜国立大学
- (1) インターネット上から利用者が着目する領域・観点に関する意見を抽出、意見間の論理的関係を分析する技術
- (2) 意見の時間的変遷等を分析する技術

### 3. 進捗状況

全体における位置づけとして、経済産業省の情報大航海プロジェクト、文部科学省のデータベースプロジェクトとの連携を行ってきた。今後、さらに文部科学省のWeb社会分析連携プロジェクトとの連携を拡充し、情報の分析の高度化と情報の活用の促進によって、次世代の情報検索、情報解析技術基盤の構築、超巨大データの戦略的活用といったような目標を達成していく位置づけである。

この研究開発の発端は情報通信審議会（総務大臣の諮問機関）の平成17年7月29日の「ユビキタスネットワーク社会に向けた研究開発の在り方について～UNS 戦略プログラム～」に端を発する。具体的には、玉石混交の情報の氾濫への対応、情報の信頼性や情報発信者の信用度に係る不安などのさまざまな課題を解決していくことが必要であるというところから来ている。

当初の目標は、平成22年までに膨大で真偽

不明な情報が混在するWebコンテンツから信頼性・信憑性の高い情報を容易に得ができる環境の実現、それから、Webコンテンツの違法・有害情報を効果的に検知し、削除、利用停止等の措置を可能にする。

なおかつ、その信頼できる情報を提供するため、情報同士の関連付け、情報の要約等々の情報分析技術の確立といったことを目標として始まったものである。

研究体制は、情報通信研究機構（NICT）が自ら行っている研究と同機構が民間に委託して行っている研究の2種類があるが、連携施策群に位置付けられているのは、NICTが委託を行っているほうの研究となる。これらは、相互補完し研究を進めている関係上、これまでに開催された「情報の巨大集積化と利活用基盤技術開発連携群の活動のシンポジウム」および「情報の巨大集積化と利活用基盤技術開発タスクフォース」でも2種類を発表している。

研究の中身は、データ解析アプローチによるWebコンテンツを構成するテキストの画像・音声・映像情報の分析に関する研究開発を「課題ア」とし、自然言語処理による情報相互の論理的関係、時間的変遷の分析技術の研究開発を「課題イ」として分類している。

これとNICTの自ら研究開発の情報信憑性分析システム（WISDOM）とが相互補完することによって誰もがWebコンテンツを安全、安心に利活用できるための技術を実現するという目標を達成するというものである。

#### a. 課題ア（Webコンテンツ分析技術）

Webコンテンツの分析技術には、コンテンツ分析に基づく信憑性分析、コンテンツに対する社会的な支持分析、コンテンツの発信者分析という三つの柱がある。

画像・音声・映像情報のクロスメディア的な分析を「課題ア-1」、テキスト情報の分析を「課題ア-2」というように二つに分類して

おり、いずれも民間の検索エンジンやポータルサイトに連動した形で実証実験を行って、結果を研究開発に再度フィードバックしていくというようなプロセスをとりながら、進めている。

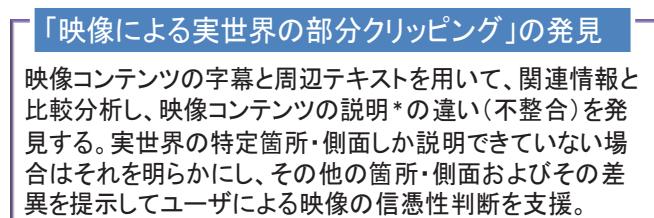
この研究の最終目標は、10以上のトピックについて、人間による評価と比較し、精度80%以上、処理時間は「画像・音声・映像情報のクロスメディア的な分析」は、現実的な処理時間、「テキスト情報の分析」は数秒から10秒以内を目標としている。

平成21年度末には、10以上のトピック（比較対象）について、人間による評価と比べて精度75%以上、処理時間に関して、「画像・音声・映像情報のクロスメディア的な分析」は、現実的な処理時間、「テキスト情報の分析」

は数秒から20秒以内の処理時間を達成する予定である。

### (1) クロスメディア型映像整合性分析

クロスメディア型映像整合性分析は、図2を例にあげると、あるテレビ局のX国大統領を対象として、分析を行うとした場合、まず、Web上の関連する同大統領の映像情報・テキスト情報等を収集し、他のテレビ局の同大統領に関するニュース映像等をすべて集め、それぞれの整合性に関して分析を行い、BTVテレビ局はX大統領とY国大統領が完全に一致したことを報道しているが、他の局では二人の関係の悪さに着目しているといったような分析結果を出す研究である。



\* : 本研究では、映像コンテンツの説明=周辺テキストとする。

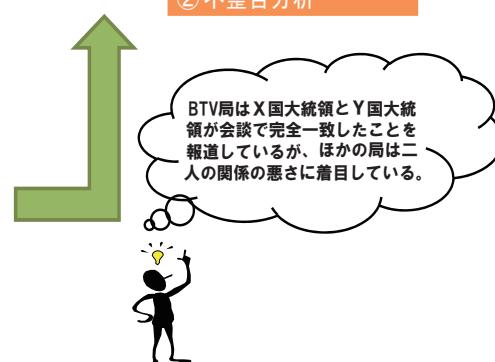
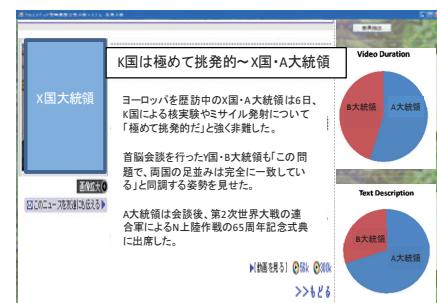
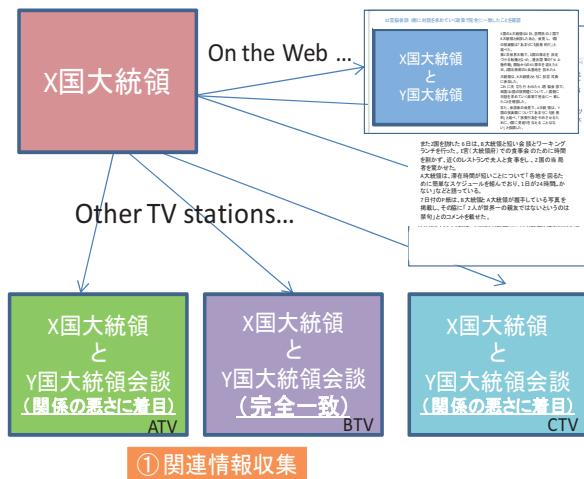


図2 クロスメディア型映像整合性分析

## Web地図画像の整合性分析

- ・方角、経路、距離の整合性分析技術の開発
- ・略地図データとオンライン地図データの対応付けにより、オブジェクトの位置関係に関する信頼性を提示

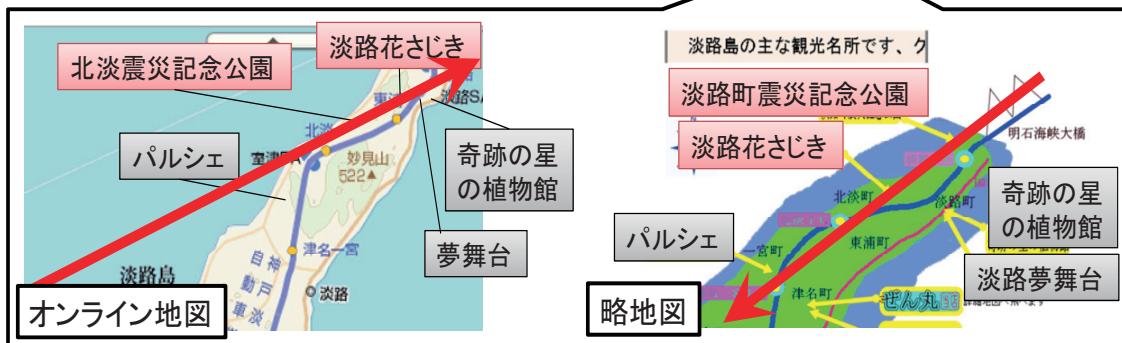


図3 実空間への要求意図に基づく地理的整合性分析

### (2) 実空間への要求意図に基づく地理的整合性分析

実空間への要求意図に基づく地理的整合性分析は、方角、経路、距離の整合性を分析するものであり、地理の2次元的な整合性の分析の開発となる。

図3を例にあげると、淡路島のWeb上の略地図を対象として、もうひとつの地図のオンラインデータとこれを照らし合わせると、淡路、花さじき、北淡町祈念公園等、位置関係を比較対照することによって、結果として、淡路花さじきと淡路町震災記念公園は位置関係が逆であるとの注意喚起をするものである。

### (3) サーチエンジンの検索データを用いたテキストの表層特徴分析

サーチエンジンの検索データを用いたテキストの表層特徴分析は、例えば、レビューや広告テキストについての信憑性の判断を行う

ものである。

図4を例にあげると、牛肉のブランドについての信憑性について判定するものだが、信憑性のあるブランドとして近江牛や松坂牛などにおいて必ず言及される属性があり、商標として登録されていれば商標のデータと照らしあわせ、ブランドが必ず言及すると思われる属性に関して、飼育方法、品質等の属性についての二部グラフを作成し、選んだ属性について多数言及されているということであれば、その分ブランドである可能性が高いという評価を行っていくものである。



図4 サーチエンジンの索引データを用いたテキストの表層的特徴分析

#### (4) Web テキスト信憑性分析：発信者のセンチメント（感情）分析

Web テキスト信憑性分析のうち、発信者のセンチメント（感情）分析は、感情表現辞書として新聞記事等から感情表現を抽出し、話題ごと、発信者ごと、組織ごとに検出するものである。ここでの感情表現とは、うれしい、怖い等の感情表現のものもあれば、日本晴れといったような間接的に表現しているものも含めた形で辞書を作っていくもので、例えば、地域ごとに八つの指標を用いて、団体 A いうものに対するセンチメントを比較対照して分析を行うといったものが考えられる。

#### (5) Web テキスト信憑性分析：発信者の熟知度分析

Web テキスト信憑性分析のうち、発信者の熟知度分析は、既に実証実験のシステムが公開されている（図5参照）。熟知領域リストおよび領域ごとに熟知している人が必ず使うであろう言葉の辞書を作っていく、過去に投稿

したエントリーの分析および熟知度スコアに基づくエントリーのランキング、各エントリーのセンチメント分析等を行っていくものである。



図5 Web テキスト信憑性分析：  
発信者の熟知度分析  
(公開中の実証実験サイト)

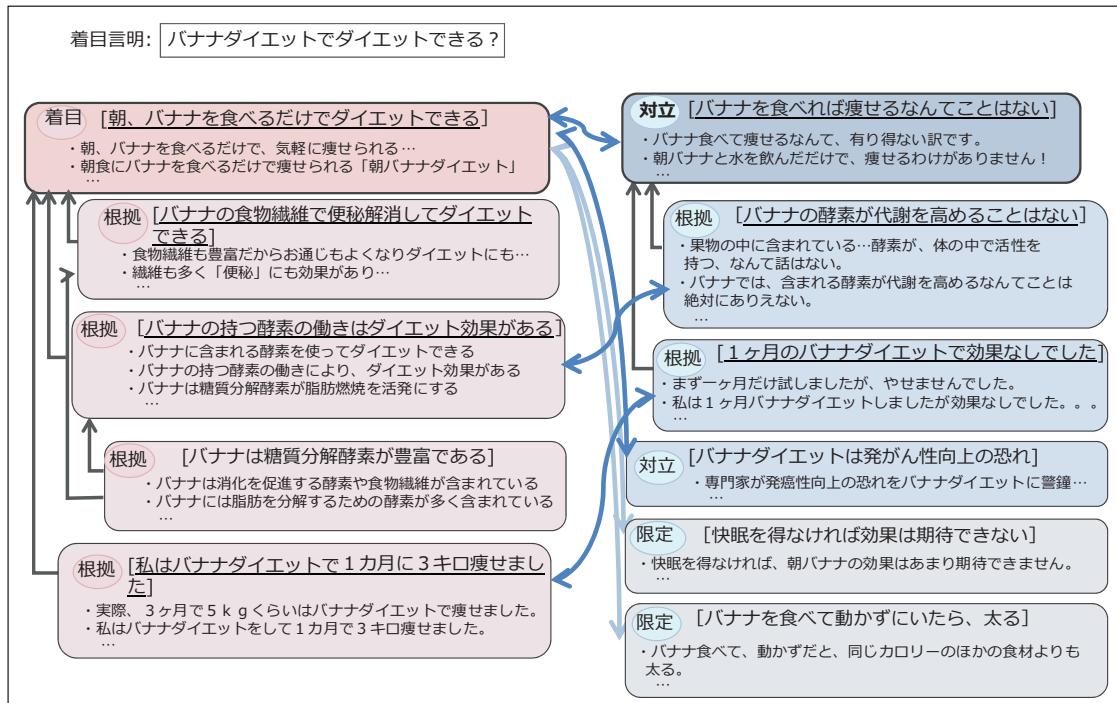


図6 最終出力イメージ (言論マップ)

### b. 課題イ (意味内容の時系列分析技術)

意味内容の時系列分析技術は、図 6 を例にあげると、朝バナナを食べることによってダイエット効果が得られるという情報が正しいのかどうか、信じていいものかどうかということで、第1に、情報を論理空間に位置づけ、朝バナナに関する評価表現などをを集め、それを反対意見になるもの、根拠となるようなものに分類し、マップを作る。

第2に、時間軸に位置づけることで、どのような経緯でこういったものが出てきたか、あるいは今でもこれは有効なのかということを時系列的に分析する。

第3に、言論空間を再構成し、反対意見、賛成意見等の全体意見のどこに注目していくべき良いかということを分析する。

最終的にはそれらを総合出力して、気になる情報の言論空間での位置づけを分析し、俯瞰的要約をユーザにレポートすることにより、広い視野から偏りのない判断を支援していくというものである。

課題イ (意味内容の時系列分析技術) の最終目標は、言論空間に位置づけるということについて、例えばダイエット等の特定の 10 トピックについては論理的関係の認識精度 80%、辞書については主に動詞に関する事象間関係知識約 50 万関係を構築する。実体間知識約 2 百万関係を精度 90%で獲得する。また、主に名詞に関する言論空間の再構成については、同様に 10 トピックにおいて 80%の精度、要約文書においては人手要約との比較で精度 80%、3 分程度で分析結果を提示する。

時間軸への位置づけについても、適切度 80%の抽出結果を 3 分程度で提示、有効期限切れ判定についても、精度 80%の言論空間関係等の目標を掲げている。

なお、平成 21 年度は統合システムを用いて実証実験を行うという目標を達成する予定である。

言論マップについては、バナナダイエットについて、バナナを食べるとダイエットできるというのに対し、根拠となる言論をピック

アップし、そこにマップを位置づけていく一方で、対立する言論をピックアップしマップ上に位置づけ、さらに対立する言論の根拠となる言論、正しいが正しさは限定的である言論などをマップ上に位置づけていく。

次に言論空間の再構成であるが、最終出力のイメージ（図7参照）ということで、俯瞰的要約を行った上で、マップで位置づけてい

った各論点の主張を並べ、対立点はどこかということにフォーカスしていく。

さらに対立点にフォーカスした上で、対立言論を擬似対立とそれ以外に分けるという作業を行う。擬似対立というのは表面的には対立しているが、内容を精査すると、必ずしも対立しているとは言えないものである。

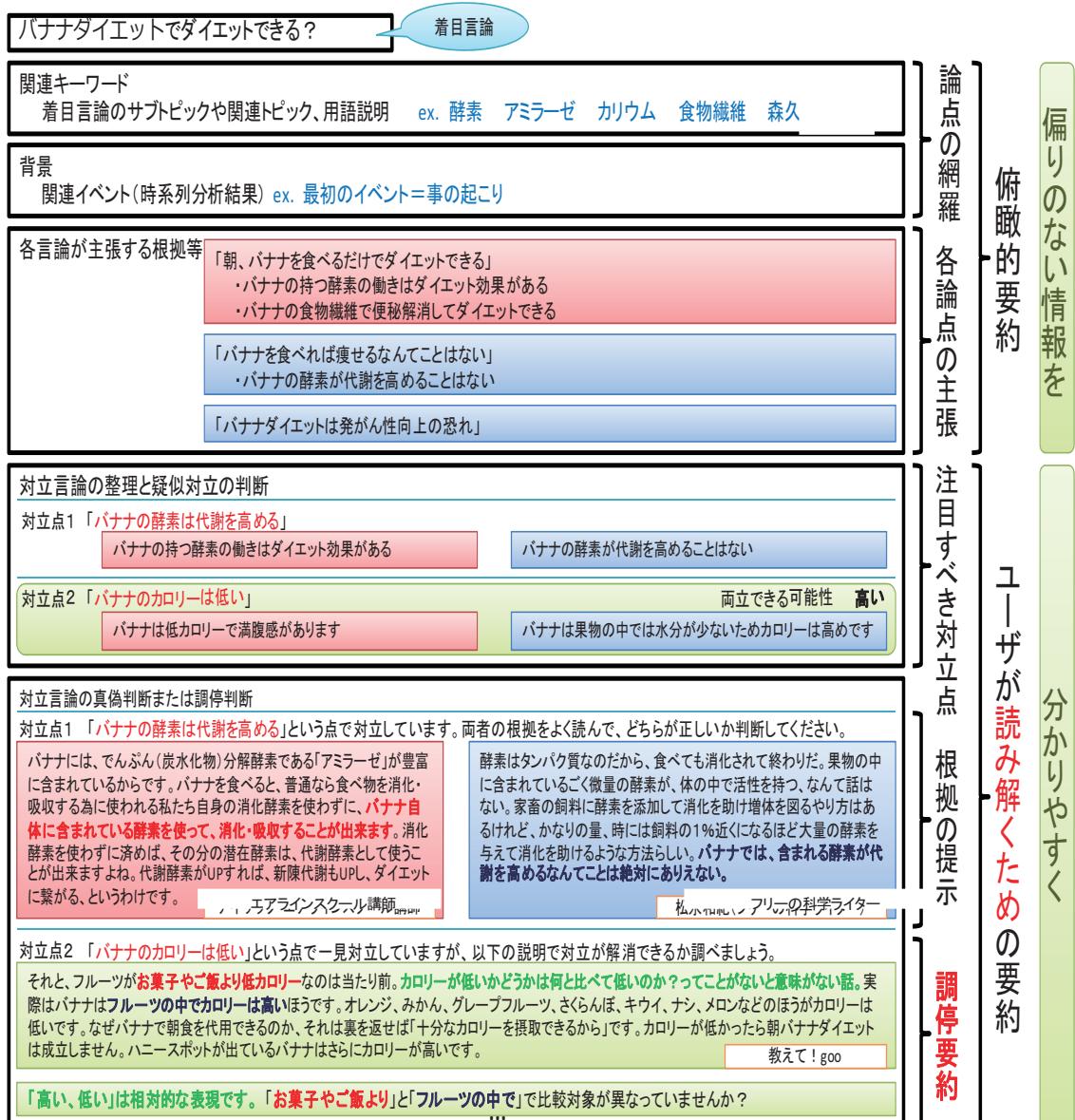


図7 最終出力イメージ（言論空間の再構成）

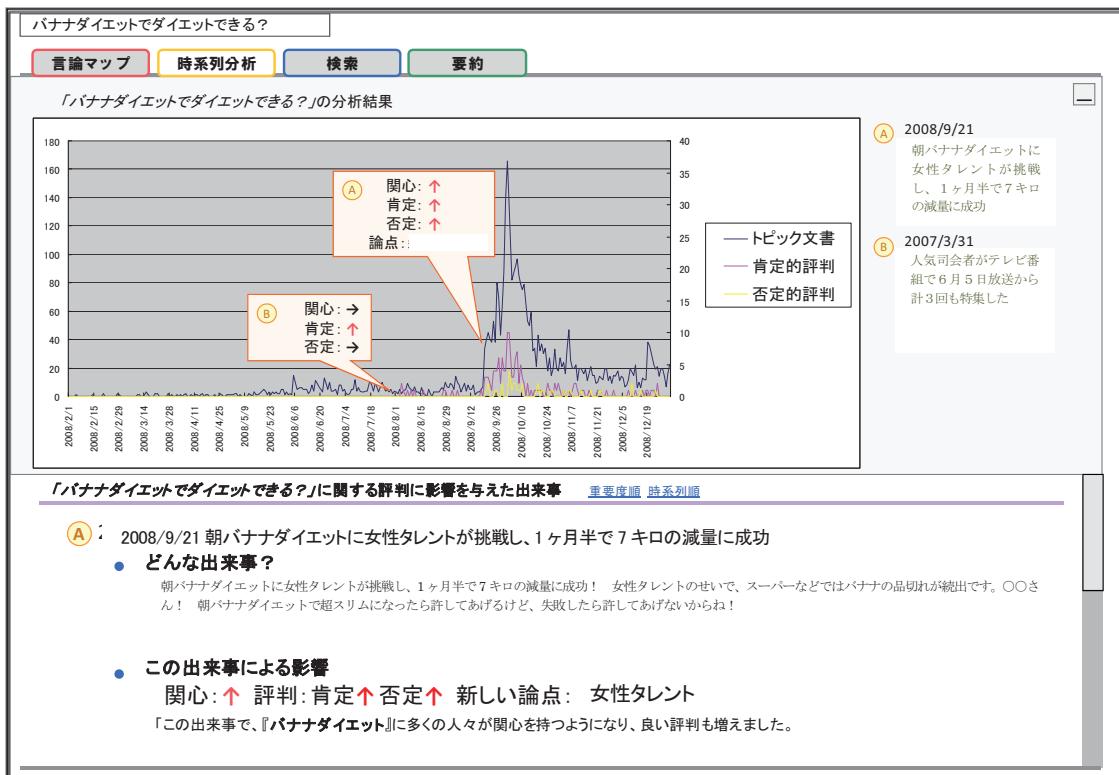


図 8 変化要因抽出の結果イメージ

つまり、一見対立しているように見えるが、観点が違うだけではないかというのが擬似対立である。

図 7 では、対立点 1 と対立点 2 があるが、対立点 2 のところで、バナナのカロリーは低い、もうひとつがバナナは果物の中では水分が少ないのでカロリーは高めという言論があるが、これらは擬似対立ではないかとして、それらを対立しない形でまとめていくこと、要約することができないかというのが調停要約を行っていく作業である。すなわち、バナナはフルーツの中ではカロリーは高いが、一般的にはカロリーは低いという形で、一つにまとめていくというような作業を行っていく試みである。

最後に言論の時間軸への位置づけについて説明する。これは三つの指標、すなわち関心、肯定、否定が急激に変化した時点はどこかを

時系列的に調べていき、急激に増えた、あるいは急激に減った原因を探るというものである。

図 8において、グラフ上大きく変化したところに何があったのかに着目すると、ここではテレビ局で有名なタレントが、1ヶ月 7Kg の減量に成功したというのを例にした。非常に大きな影響を与えたものの分析を行うものである。

統合システムは、課題イの各々三つの研究開発は現在、分散して実施されているが、今年度末までに、これらのエンジンを統合するシステムを開発し、図 9 のように実証実験を行う。

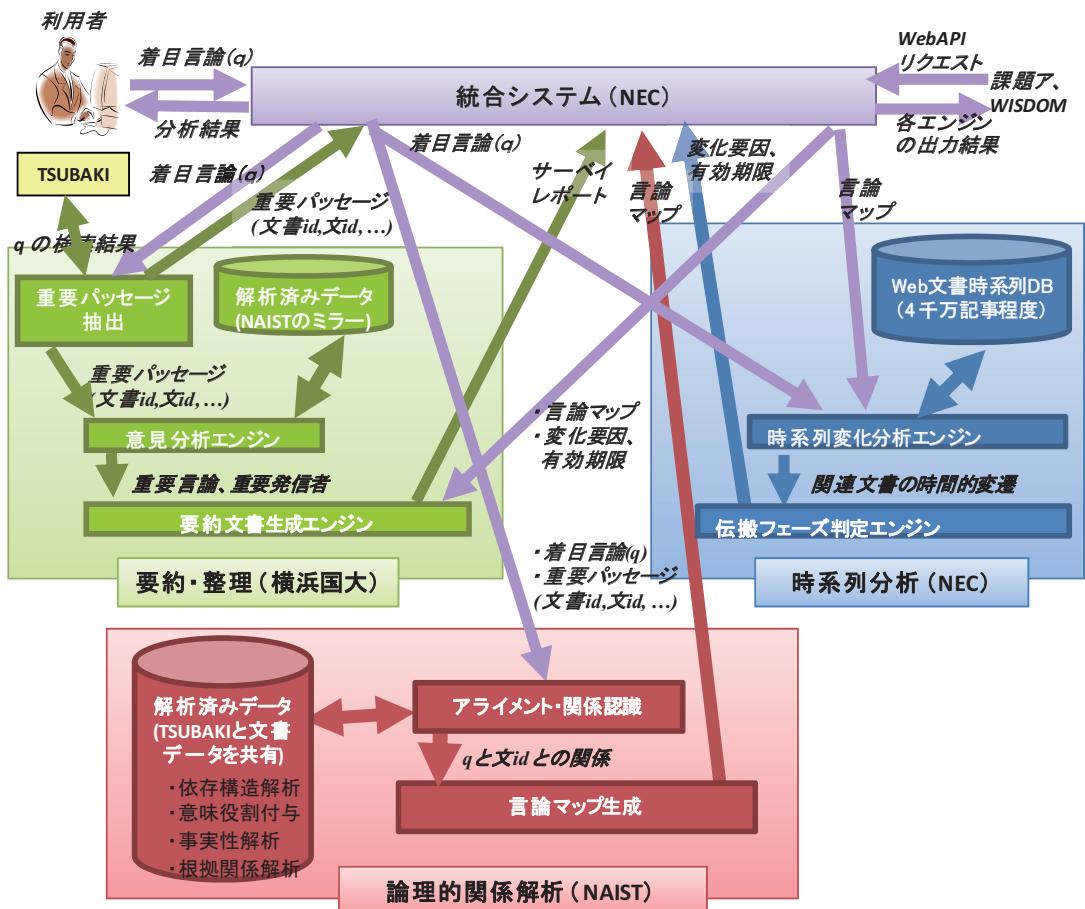


図9 実証実験イメージ

### c. 情報信憑性分析システム (WISDOM) と の相互補完

#### (1) 課題ア

WISDOM は自前の Web 検索エンジンを用いた詳細な自然言語処理による信憑性検証を行うシステムであり、一方、課題アで開発しているものは、既存の各種 Web 検索エンジンを用いた軽量な信憑性検証システムである。

また、WISDOM は日本語コンテンツのみを対象としているが、課題アで開発するものは日本語コンテンツに加えて英語コンテンツも対象とするものである。

両者ともに、機能や対象データは異なるが、双方ともに、コンテンツ自身の分析、コンテンツの社会的支持度の分析、コンテンツの著作者分析を行う点では同一方向のものである。

したがって、両システムを相互補完的に用

いて情報信憑性判断支援ができるような形で連携させる。具体的には、軽量な課題アのシステムで情報の検索及び検索結果の軽量な信憑性検証を行い、さらに詳細な分析を行う場合に WISDOM が連携して動作するといったシステム連携を想定している。

#### (2) 課題イ

課題イで開発している実験システムの分析機能を WISDOM の追加機能として導入する。

これにより、利用者は WISDOM の従来機能を用いて、信頼性判断を行いたいテキスト情報を関係して Web 上に書かれている情報の大まかな傾向を調べ、さらに重要なポイントを課題イの機能で深く調べるシームレスな利用が可能となるシステム連携を想定している。

#### 4. 関連技術の施策

関連する技術として、平成 21 年度から 3 カ年計画で、NICT から（株）KDDI 研究所に委託している研究で、インターネット上の違法・有害情報を検出する技術の研究開発がある（図 10）。

これは平成 20 年 6 月に起きた秋葉原での連続殺傷事件がきっかけとなり、インターネット上の違法・有害情報を効率的に探索するの

が非常に重要ではないかということで、信憑性検証技術とも関係するが、現在は人手に頼る部分が大きい違法・有害情報をインターネット上から検出する技術を開発する。

単に検出するだけでなく、例えば薬物や銃器などは隠語を使っているような場合もあり、隠語を含めて検出できる技術の研究開発を現在、進めている。

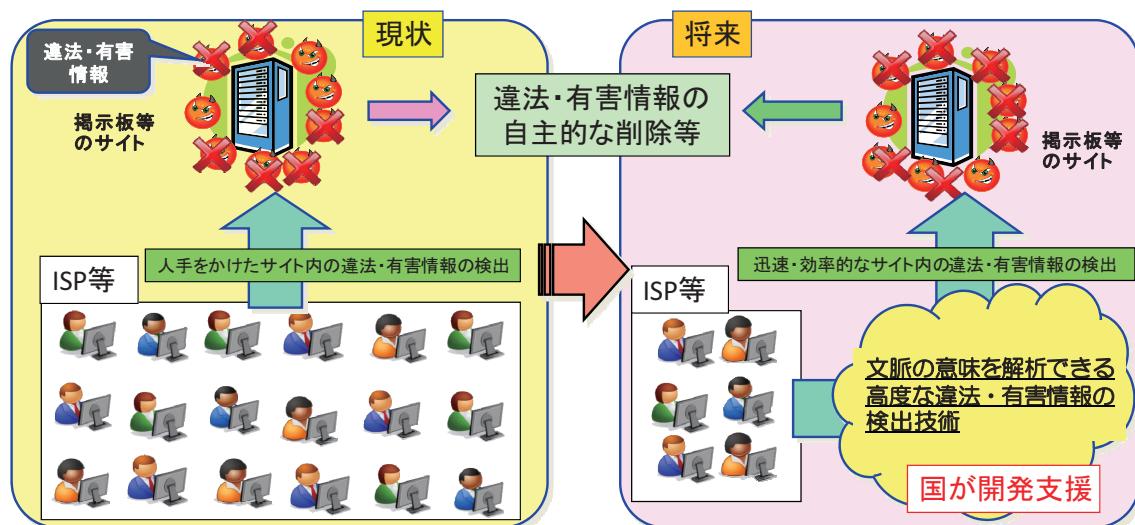


図 10 インターネット上の違法・有害情報の検出技術

### 3.3 革新的実行原理に基づく 超高性能データベース基盤ソフトウェアの開発

文部科学省研究振興局情報課 情報科学技術研究推進官  
後藤 祐介

#### 概要

情報の巨大集積化と利活用基盤技術開発連携群における取り組みとして、平成19年度より「革新的実行原理に基づく超高性能データベース基盤ソフトウェアの開発」を行ってきた。連携施策群において本施策は、情報大航海プロジェクトや信憑性検証プロジェクトが扱う大量かつ多種な情報を効率的に扱うための技術基盤として位置づけられている。

#### 1. 背景・目的

今日の情報化社会におけるデータベースの役割は、非常に重要なものとなっている。IT利用の進展に伴い、大量の情報を適切に管理し、その中から有益な情報を見つけ出すことが必要である。さらに、近年では情報量は爆発的に増大し続けており、今後はさらに飛躍的に情報量が増大することが予想されている。このような状況から、増え続ける膨大な情報を高速に処理し、戦略的に活用していくことが、我が国が新規産業・高付加価値産業を継続的に創出し続け、国際競争力を維持・強化するとともに、さらには、国民生活の安全・安心を確保するためにも不可欠な課題である。

#### 2. 研究内容

大量の情報の適切な管理と処理を担うデータベースシステムの性能は、プロセッサなどハードウェア技術の高性能化等により向上をしてきたが、爆発的に増大を続ける情報をデータベースで高速に管理するためには、ハードウェア技術の高性能化だけでは十分でない。このことから、データベース実行原理の革新によって、すなわち核となるソフトウェアの抜本的な変革によってデータベースシステムの性能ブレークスルーを実現することが、これらの課題を解決するための有効な手段である。

あると考えられる。この革新的なデータベースシステムを実現するために、本施策では、非順序型実行原理に基づく超高性能データベースエンジンの開発を実施してきた。

研究は、東京大学生産技術研究所の喜連川教授を研究代表者とし、東京大学生産技術研究所を研究拠点として、東京大学と日立製作所の产学研連携体制により実施している。

図1に従来型のデータベースエンジンと、本研究で開発を進めている非順序型のデータベースエンジンの差異を概念的に示す。従来型のデータベースエンジンでは、ある問い合わせ処理のためには、逐次的にストレージ(ディスクドライブなどのデータが記憶されたメディア)へのアクセスを繰り返す。つまり、あるデータのアクセス要求を行ってから、その要求に対する応答を得て、次のデータのアクセス要求を行うというように処理が進行し、順番待ちの必要がある。一方、非順序型データベースエンジンでは、一度に大量のストレージへのアクセス要求を行い、ストレージ内で見つかったデータから処理をしていくことができる点に特徴がある。

例えばある小売業者では、RFID(電子タグ)を用いた製造・流通・販売に跨った抜本的効率化を試みているが、データベースのサイズは優にペタバイト(10の15乗バイト)を超えるとみられており、このように一つ一つの

レコードは小さいけれども、数が極めて多いデータベースに対しては飛躍的な性能向上を期待することができる。

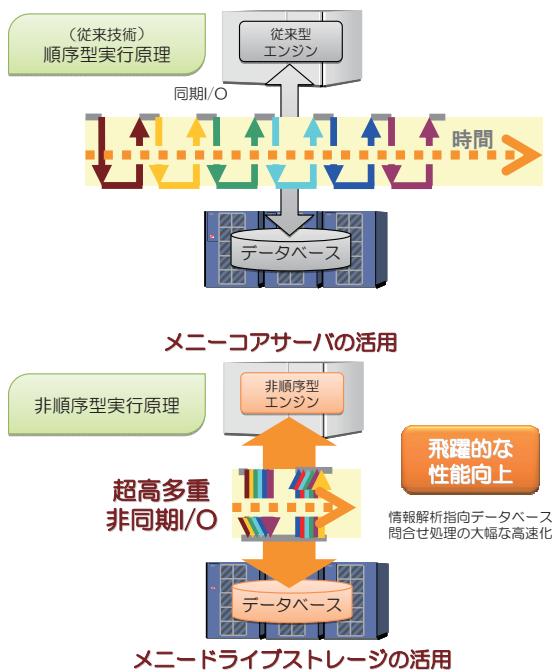


図1 非順序型データベース実行原理

### 3. 研究の目標

本研究は、平成19年度から始動しており、平成21年度中に、非順序型実行原理を一部のデータベース処理に適用した限定版の非順序型データベースエンジンを開発し、従来技術の約10倍のデータベース解析処理性能を達成することを目標としている。

また、平成23年度には、非順序型の実行原理を本格的に適用した非順序型データベースエンジンを開発し、従来技術の約100倍を達成することを目指して研究開発を行っている。

### 4. 主な進捗状況

平成19年から、これまでに連携施策群のシンポジウムや、文部科学省の情報科学技術委員会等において有識者による中間評価を行っている。中間評価においては、我が国発の新しいデータベースエンジニアーキテクチャの確立を十分に期待させるものであるとして、

高い評価を頂いている。

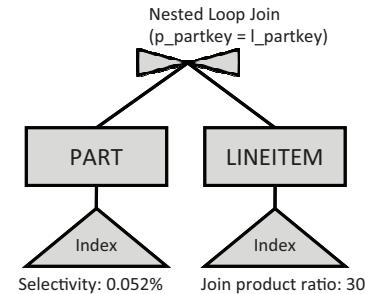
研究計画では、平成21年度中に、従来技術と比較して約10倍の性能向上を達成することを目標としていたが、平成21年12月時点での業界標準のベンチマークの一部問い合わせにおいて、商用のデータベース管理システムであるHiRDBベースの実験で約20倍、オープンソースのデータベース管理システムであるMySQLベースの実験で約30倍の性能向上を達成するなどの成果が上がっている。

### 5. 具体的な研究成果について

まず、オープンソースのデータベース管理システムであるMySQLをベースとする非順序型データベースエンジンの開発については、マルチコア環境への対応などの飛躍的な性能向上に向けた萌芽的諸課題への取り組みに注力して非順序型データベースエンジンの開発をすすめている。その性能を評価したところ、図2のとおり、ベンチマークの一部の問い合わせにおいて、順序型データベースエンジンと比べて、30倍を超える高速化を達成した。

次に、商用データベース管理システムであるHiRDBをベースとする非順序型データベースエンジンについては、むしろ商用化を視野に入れ、商用システムへの組み込み難易度を精査しながら開発を進めており、業界標準のベンチマークの複雑度の高い問い合わせにおいて、図3のとおり、約20倍の高速化を達成している。平成21年度内には、ベンチマークのうち半数の問い合わせをサポートする予定である。

TPC-Hベンチマーク  
データセット : SF=100 (100GB)  
問合せ : Q.14相当



4x Quad-core Xeon Processors  
32GB Memory (Only 1.6GB used)  
4x FC HBAs (4Gbps)  
20x 10Krpm 146GB FC HDDs (RAID-0)  
RedHat Enterprise Linux 5.3  
MySQL 5.1 InnoDB Storage Engine

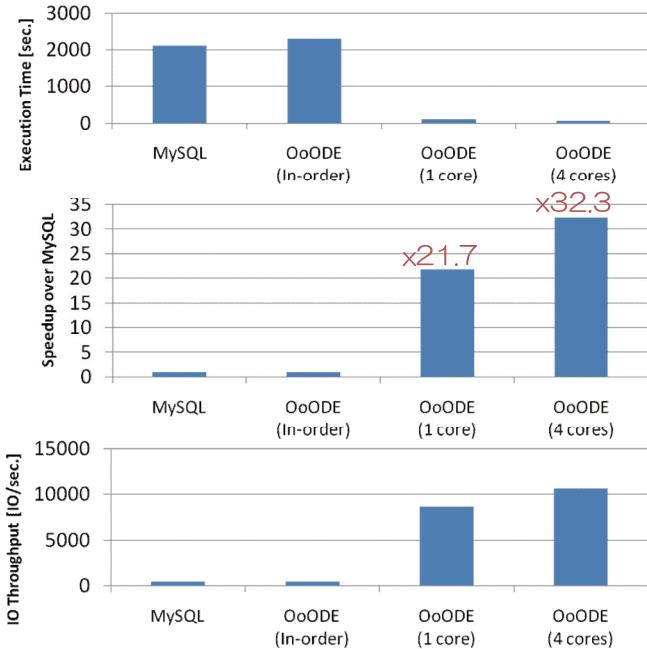
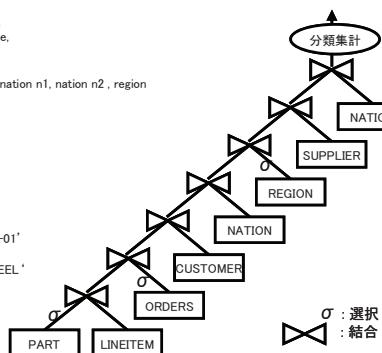


図2 MySQLにおけるマルチコア活用

#### 複雑度の高い問合せにおける性能向上

(例：指定地域における指定商品種別の2年間の市場シェア変化を分析；TPC-H Q.8)

```
SELECT o_year,
       sum(case when nation = 'BRAZIL' then volume else 0 end) / sum(volume) as mkt_share
  FROM (
    SELECT
      extract(year from o.orderdate) as o_year,
      l_extendedprice * (1-l_discount) as volume,
      n2.n.name as nation
     FROM
      part, supplier, lineitem, orders, customer, nation n1, nation n2, region
    WHERE
      p_partkey = l_partkey
      and s_suppkey = l_suppkey
      and l_orderkey = o_orderkey
      and o_custkey = c_custkey
      and c_nationkey = n1.nationkey
      and n1.n_regionkey = r.regionkey
      and r.name = 'AMERICA'
      and s_nationkey = n2.nationkey
      and o_orderdate between date '1995-01-01'
                        and date '1996-12-31'
      and p_type = 'ECONOMY ANODIZED STEEL'
      and p_size < 3
   ) as all_nations
 GROUP BY o_year
 ORDER BY o_year;
```



Intel Xeon 2.66GHz, 32GB,  
HDD (100Krpm) 20台, TPC-H SF=1000

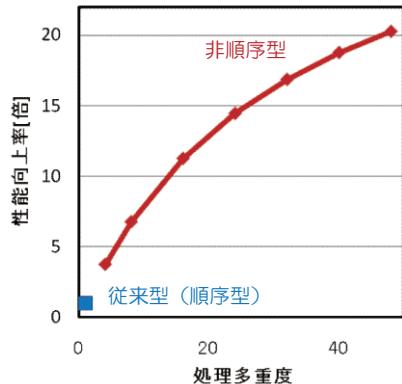


図3 HiRDBベースの性能向上

なお、非順序型実行原理の下では、ソフトウェアは非決定的な挙動を基本とすることから、ソフトウェア動作の正当性の検証や性能デバッグは困難を極め、従来型のソフトウェア開発支援ツールでは十分でない。本研究では、非順序型データベースエンジン、OSならびにストレージの挙動を精緻に観測するモニタリング機構を開発している。入

出力機構の解析によって、特異挙動の解明や、タスク間の長い排他待ちの解析を可能とするなど、困難なデータベースエンジン開発の効率化を実現している（図4）。

- 【データベースエンジン】
  - ・エンジン内部処理遷移
  - ・CPU処理内訳
  - ・スレッド処理、I/O発行挙動
  - ・リソース使用状況
  
- 【OSカーネル】
  - ・OS内部処理遷移
  - ・I/Oスケジューラ処理時間
  - ・I/O管理キュー等のリソース使用量、及び滞留時間
  - ・I/Oスループット
  - ・CPUリソースの使用状況
  - ・プロセススケジューリング状況
  - ・SMP環境におけるOS内ロック状況
  
- 【ストレージコントローラ】
  - ・ストレージ内部処理遷移
  - ・キュー、CPU、HDD等のリソース使用状況
  - ・I/O応答時間、スループット等の統計

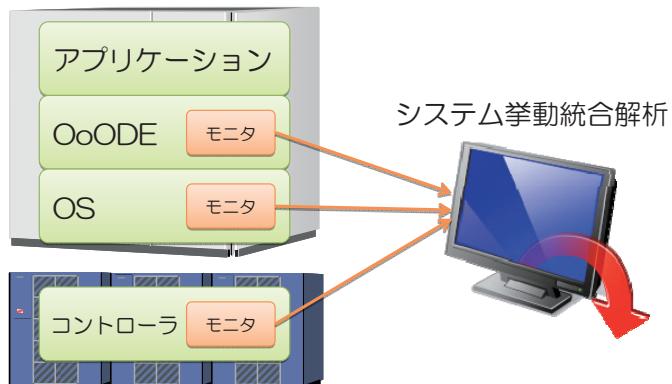


図4 挙動モニタリング機構

## 6. 今後の展望について

本研究の成果が具体化することにより、様々な分野で、これまで以上にデータベースを戦略的に活用することが可能となることが考えられる。冒頭でも述べたとおり、世界に先駆けた非順序型データベース実行原理を確立することによるデータベース管理システムなどの基盤ソフトウェア産業の国際競争力が向上し、我が国産業の国際競争力を強化するなどの効果が期待される。

例えば、トレーサビリティやセンサ技術等の分野で利用することにより、従来では情報量が多くて処理できなかったような情報を有効活用することが可能となり、情報科学技術による斬新な社会サービス応用が創出され、国民生活の安全・安心の向上にも寄与することも期待できる。

情報科学技術を活用し、社会の様々な複雑で困難な現実問題を解決するには、大規模・大量のデータの蓄積・利活用を可能とする研究が不可欠である。また、センサ情報の高度な活用には匿名化等のプライバシ配慮技術が組み込まれることは言うまでもないが、社会への普及の観点からは技術面だけの取り組みだけでなく、社会制度・法制度の面からの検討も重要である。今後はセンサ情報が情報爆発の中心となり、多様なセンサ情報から実世

界の情報を収集し、これに高度な解析を施すことにより社会システムの抜本的な効率化を目指す CPS (サイバー・フィジカル・システム) サービスが構築されていくと見られ、本研究による革新的な超高性能データベース基盤ソフトウェアがその重要な一基盤技術となるであろう。

なお、本研究開発は、平成 23 年度までの予定であったが、平成 21 年、総合科学技術会議で決定された最先端研究開発支援プログラムで、本研究と同旨の喜連川教授の研究課題が選定された。このため、本研究については、平成 22 年度以降は、同プログラムへ発展的に一本化することとなっている。

### 3.4 多メディア Web 解析基盤の構築及び社会分析ソフトウェアの開発

文部科学省研究振興局情報課 情報科学技術研究推進官

後藤 祐介

#### 概要

「Web 社会分析基盤ソフトウェアの研究開発」は、平成21年度より新規に立ち上がったもので、公募の結果、国立情報学研究所を中心とする体制により実施をすることが決定し、8月に事業を開始したところである。従って、事業の概要を中心に今後の研究開発計画等について述べる。

本施策の連携施策群における位置づけは、情報大航海プロジェクトの一部成果の活用や信憑性検証プロジェクトと連携し、蓄積した Web 情報の高度な分析を行う基盤の構築として位置づけられている。

#### 1.背景・目的

本施策は、社会学、言語学、リスク管理、マーケティング等多様な社会分析ニーズに応じるために、膨大な多メディア Web 情報を収集、蓄積し多様な解析を可能とする多メディア Web 情報解析基盤の構築を目的としている。

爆発的な増加を続ける Web 情報は、画像・映像等への多メディア化が急速に進むと同時に、放送映像等の実世界情報と相互に及ぼし合う影響も拡大し続けている。

そこで、こういった Web 情報をアーカイブし、時系列で分析する技術を開発することで、人類社会の観測・調査・解析が可能となり、新たな価値を創出する可能性がある。

#### 2. 研究目標・開発内容

目標としては、研究期間中に多メディア Web 解析のための要素技術ならびに基盤技術の実用規模での実装を行い、多メディア統合処理プラットフォーム上で統合し、社会分析ソフトウェアの実証実験を行うところまでを目指している。

そのための研究開発内容は、一つ目に、多メディア Web 解析のための要素技術に関する研究、二つ目に多メディア Web 収集、蓄積および高速処理に関する研究、三つ目にその両者を統合する研究、四つ目として社会分析のためのアプリケーションについて研究を行うこととしている。

これらの研究開発内容の構成ならびに関連を図 1 に示している。

本施策の特徴として、

- ・多メディア Web の大規模・時系列的な収集・蓄積による社会分析の情報源確保
- ・多メディア Web 解析による特定の話題の出現の検出
- ・話題の発生・発展・反響等の時系列的追跡

これらの解析結果ならびにその可視化等

が可能になることが挙げられる。

#### 3. 期待される効果

学術面では、社会学、言語学、経営学等、多様な学術分野におけるこれまでにない新たな学術調査手法の創出が期待できる。

産業面では、従来手法では十分把握できなかった国民生活の実態を捉えることやテレビ番組や Web 上の動画を通じて企業やコンテンツ制作者から発信された情報がどのように視聴者・消費者に影響を与えたかを詳細に解析することが可能となり、新規ビジネスの創出や高度なマーケティング手法の現出が期待できる。

また、本研究開発は国内外において他に例を見ず、独自性が高く、国際競争力が期待できる。

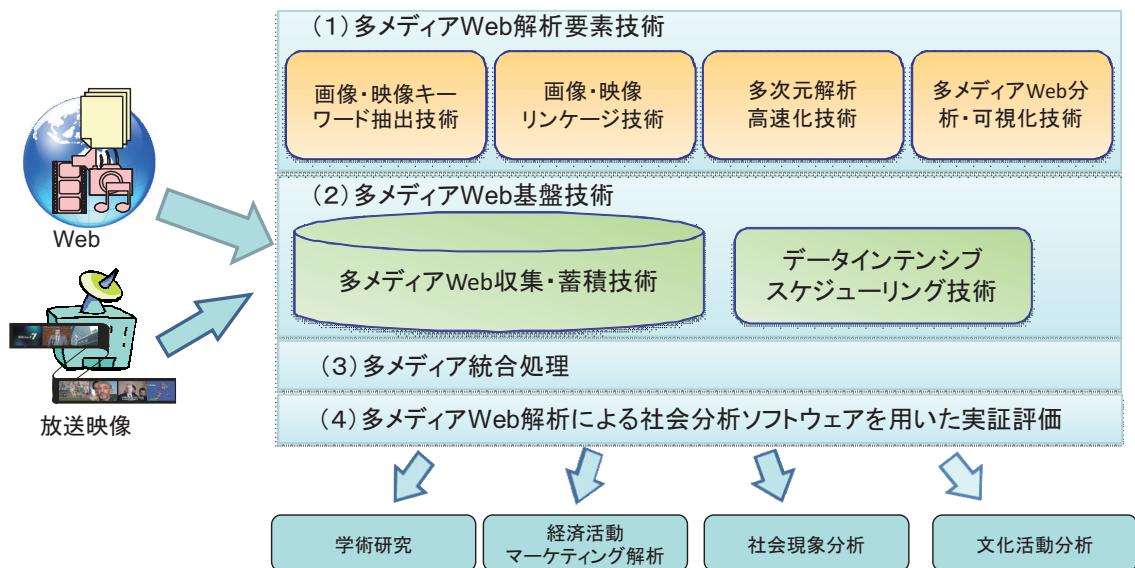


図1 「Web 社会分析基ソフトウェア」の研究開発内容

#### 4. 研究実施体制

国立情報学研究所の佐藤真一教授を研究代表者として、国立情報学研究所、東京大学、早稲田大学の三機関を研究拠点として実施している。

また、成果を用いた学術調査への利用や実サービス実施に対する要請、法制度的問題点等に関する検討を行うため、Web アーカイブ構築活用助言委員会を設置し、研究開発の方針を定めていく予定としている。

#### 5. 研究開発計画

研究開発は平成 21 年度から平成 24 年度までの計画である。平成 21 年度および平成 22 年度は、「多メディア Web 解析要素技術及び基盤技術」の基本設計、社会分析ソフトウェアのプロトタイプ実装を予定している。

平成 23 年度以降は、これらの技術の統合を行い、最終年度には社会分析ソフトウェアの実用規模での実証実験を行う予定としている。

#### 6. 今後の展望について

我が国を取り巻く世界情勢は、第 3 期科学技術基本計画が策定された平成 18 年以降、世界同時不況の発生に始まり、地球温暖化問題や環境問題など複雑かつ困難な状況が一層顕在化しており、これらの社会的問題の解決に向け、情報科学技術への期待はますます増大しつつある。

特に、科学技術の発展のみならず、その成果を還元して、社会全体のイノベーションにつなげていく視点が重要であり、その観点から今後の社会システムを構築する上で重要な基盤となっている情報科学技術の果たすべき役割は大きなものと考えられる。

文部科学省としては、我が国における情報科学技術に関する研究開発の着実な進展に向けた支援を行っているところであり、今後ともその推進に努力して参りたい。

### 3.5 センサ情報の社会利用のためのコンテンツ化

京都大学学術情報メディアセンター センター長 教授  
美濃 導彦

#### 概要

ユビキタスセンサネットワーク（以下「USN」という。）から得られるセンサ情報を公開して、Web のように誰もが自由に利用できる「センシング Web」を実現することを目指し、研究開発を行った。USN をセンシング Web に進化させるには、センサ情報に含まれるプライバシ情報を取り除いて公開するプライバシ情報管理技術、センサ情報を統一した形式により表現し提供するセンサ情報共有技術、得られた実世界の観測情報をコンテンツとして提示する観測型コンテンツ提示技術の三つが必要であるという考えに基づき、それぞれの技術開発をサブテーマとし、研究を推進した。開発した技術を統合し、その社会的な有効性、受容性を検討するために、公共環境での実証実験を実施した。

#### 1. センシング Web

##### a. 研究目的

近年、道路や駅構内、建物や街中など、我々が生活している実世界環境にカメラやマイクを含む様々なセンサを設置し、それらをネットワークで結んだ USN を構築する動きが社会で活発となっている。このような USN から得られる情報は、実世界を直接観測して得られる生のセンサ情報であることから、Web から得られる情報のように人が作成した編集型のコンテンツとは対照的な観測型のコンテンツとして社会で利活用できる可能性がある。

今日の Web は、人間のあらゆる活動に必須の情報基盤となるまでに発展を遂げている。このような発展は、ARPA ネットや USE ネット等、それぞれ特定の用途のために構築された個々のネットワーク同士が相互接続され、同一のプロトコル (TCP/IP) でデータをやり取り可能なインターネットが出現したことに始まる。しかし、Web 情報の社会的利活用がここまで進んだのは、さらに、1) WWW の開発により、同一のプロトコル (HTTP) や情報記述形式 (HTML) により各サイトの情報にアクセスすることが可能となり、サイト群の情報が地球規模で実世界の情報を集積した巨大な一

つのコンテンツとして機能するようになったこと、2) 検索エンジンやポータルの開発により、そのような巨大なコンテンツに含まれる情報を利用者に効果的に提示することが可能となり、コンテンツ閲覧のための環境が整備されたこと、の 2 点による。

これに対し、現在の USN は、いずれも、自治体や企業等の組織が、道路交通管制やビル管理、人流調査といったその組織独自の用途に利用することを目的として、組織自らがセンサの種類・配置の設計や設置を行うと共に、得られたデータも組織内のみで利用するという運用が行われている。これは、Web でいえば ARPA ネットや USE ネット等が個別に構築されていた時期に相当する。このように目的指向で設計されたセンサネットワークを自設し、そこから得られる情報のみを個々の USN 毎に利用するという運用方法は、適用分野や有用性が明確で、情報管理も容易である反面、想定された目的以外への再利用性が低く、利用範囲も限られるため、Web と比べると、一般市民への貢献は極めて限定的なものに留まらざるを得ない。

そこで本研究課題では、上のような USN がごく近い将来に社会全体に普及し、世界各地

に様々な用途の USN が多数設置されている状況が出現するという予想の下に、その次の段階として、USN から得られるセンサ情報を Web のように誰もが自由に利用できる仕組みを実現することを目指す。これにより、実世界の生の観測情報を地球規模で提供する人類共通の観測型実世界コンテンツが出現することになる。このようなコンテンツを本研究では“センシング Web”と呼ぶ。

USN をセンシング Web へと進化させるためには、Web における WWW や検索エンジン・ポータルに対応する技術を開発する必要がある。Web とセンシング Web を比較すると、どちらも情報伝送、コンテンツ化、コンテンツ構造化・提示の三技術から成る。一方、両者は情報の形態の点で異なり、Web の情報は人手で入力される編集型の情報であり、テキストデータを中心とするのに対し、センシング Web の情報は、センサにより得られる観測型の情報であり、パターンデータとなる。また、観測型の情報は、編集型の情報とは異なり、被観測者のプライバシ情報が含まれているため、利用者に応じてこれを保護するような仕組みも必要となる。

このうちプライバシ情報の管理のためにはパターン処理技術、センサ情報の共有基盤の確立のためには情報システム技術、観測型コンテンツの提示のためにはメディア応用技術がそれぞれ重要な役割を果たすことから、これら三つの研究分野の研究者を結集し、緊密な連携の下にセンサ情報のコンテンツ化と構造化・提示およびプライバシ情報管理のための技術を開発する。この三つの研究分野を図 1 に示す。

開発した技術が実社会で役立つためには、個々の技術の精度が高いだけでなく、それらを統合したシステムが全体として有効に機能することが必要である。さらに、社会的な利用を考えると、「センシング Web」のコンセプトを一般市民が理解し、受容することが重

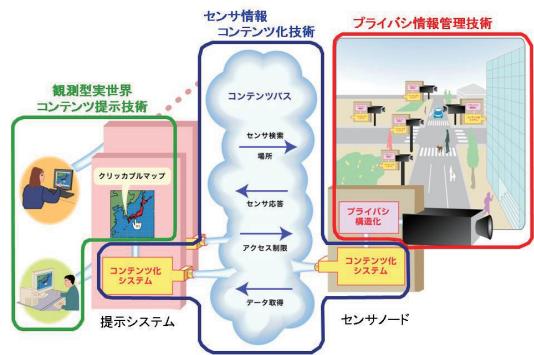


図 1 三つの研究課題

要である。そこで、開発した技術を統合したシステムを構築し、実際の公共環境に適用する実証実験を行った。また、実証実験において「センシング Web」のコンセプトを来場者に説明し、主にプライバシ保護の観点から技術の受容性を明かにするアンケート調査も行った。これらを通じて、「センシング Web」実現に向けた可能性と課題を明らかにすることを目的とする。

## b. 研究成果

センシング Web を実現するには、プライバシ情報管理技術、センサ情報共有技術、観測型コンテンツ提示技術の三つが必要であるという考えに基づき、それぞれの技術開発をサブテーマとし、研究を推進した。それぞれの詳細については、対応する章で述べ、ここでは、開発した技術を評価するために行った実証実験について述べる。

### (1) 実証実験

本研究課題の実証実験の主目的は、開発した技術の社会からの受容性を評価することである。具体的には、プライバシ問題と有用性という二つの側面について明らかにする必要がある。これらを明らかにするための実証実験会場として、駅・大学・商店街・ショッピングモール等が選択肢として挙がった。一方で、実証実験遂行上、以下のような要件も考慮する必要がある。

- ・管理者と無関係な人間が多数出入りする

- ・機器盗難のリスクが少ない
- ・利用者全員にセンサの設置を周知できる

上述の候補のうち、ショッピングモールがこれらの要件を最もよく満たすと考え、京都市内の新風館に会場利用を打診し、交渉を行った。交渉の過程では、新風館のみならず、その施設所有者である NTT 都市開発株式会社や、新風館内各テナントへの説明も行った。その結果、平成20年5月に会場利用の了承を得た。新風館は、三階建てのショッピングモールで、吹き抜けになった中庭をロの字型に廊下が取り囲み、その廊下沿いにテナントが入るという構造である。中庭には屋根はなく、廊下部分は屋外光が差し込む半屋外施設である。

センサの設置位置や設置方法の決定については、新風館に確認を取り利用客からの印象に十分に配慮しながら進めた。また、構築したシステムで取得されるデータの取り扱いについては、弁護士の指導のもとプライバシーポリシーを定め、それを順守する体制を定めた。実証実験システムを構築するにあたり、以下の情報サービスを実施することを考えた。

- ・「変身カメラ」：これは、カメラからの映像情報に対して、映像中の人物像を消去し、人物の位置のみを表すキャラクタに「変身」させて表示する情報サービスで、本研究課題によるプライバシ保護技術を一般に分かりやすく示すものである。
- ・「にぎわいマップ」：このマップは、新風館の平面地図にセンサから取得された館内の混雑情報をリアルタイムに表示する情報サービスで、センサ情報を統計情報化することでプライバシ保護を行う技術を示すものである。
- ・「デジタルジオラマ」：3 次元モデルにより再現した新風館のウォークスルーを実現する情報サービスで、実世界のセンサ情報を観測位置と結びつけて可視化する提示技術、および利用者と被写体との関係により

プライバシ情報の公開程度を制御する技術を示すものである。

- ・「シースルービジョン」：これは、利用者視点に応じた映像提示を実現する情報サービスで、利用者から見えない範囲の様子を、センサ情報で補うことで、より豊富な情報を利用者に提供する技術を示すものである。

これらの情報サービスを実施するにあたり、以下の 3 種類のセンサを選定した。

#### (a) カメラ

次の 3 種類のカメラを使用した。

- ・一般的な監視カメラ
- ・ワイドダイナミックレンジカメラ
- ・高解像度カメラ

一般的な監視カメラは、各階への人の出入りを撮影できるよう、主に新風館 1、2、3 階の階段付近 14 箇所に設置した。ワイドダイナミックレンジカメラは、新風館 3 階全体を撮影できるよう、20 箇所に設置した。高解像度カメラは、中庭の人の動きが撮影できるよう、2 箇所に設置した。

#### (b) RFID タグリーダ

RFID タグリーダは、微弱電波を発信するアクティブ型 RFID タグが一定範囲内に来ると、その ID 情報や電波強度などを取得することができるデバイスである。この RFID タグリーダを、新風館 3 階の店舗前の廊下 11 箇所に設置した。

#### (c) 赤外線センサ

赤外線センサは、その設置位置を人物が通過したかどうかを観測するセンサである。この赤外線センサを新風館 3 階の階段等ゲート部分 6 箇所に設置した。

以上のセンサは、いずれも有線 LAN で接続され、別途用意したセンサノード用 PC でプライバシ情報フィルタリング処理が施される。ここで処理されたデータは、ネットワークに

公開・共有され、他の情報サービス等に利活用される。センサの設置等、システム構築を行い、平成21年7月より実証実験を開始した。実証実験にあたっては、一般市民への広報と一般市民からのフィードバックが重要な目的であるので、新風館2階の一室にプロジェクトルームを設け、本研究課題の概要や実験内容に関する紹介を常設展示した。実証実験開始前には、本研究課題に関する記者発表を行った。また、実証実験の期間中の平成21年7月、9月、11月に、一般来場者を対象に本研究課題で開発した技術のデモンストレーションを行うイベントを開催するなど、広報活動を行った。その結果、本研究課題に関して、新聞、テレビに14件取り上げられた。

## (2) アンケート調査

### (a) アンケートの概要

実証実験の期間中の平成21年7月、9月、11月に開催した一般来場者向けのイベントにおいて、本研究課題の技術に関するアンケート調査を行った。

アンケートの目的は、以下の3点である。

- ・プライバシ保護技術に対する意識調査
- ・プライバシ保護技術の有効性評価
- ・各情報サービスの受容度評価

本研究課題では、センサ情報に対してプライバシ保護技術を適用し、個人を特定できない形で公開するというアプローチをとっている。「変身カメラ」の情報サービスを通じて、このアプローチが理解され、社会に受け入れ可能であるかを調査した。

また、利用者のプライバシ保護と利用者が受けられるサービスの有用性との間には、トレードオフの関係があると考えられる。このトレードオフについて、情報サービス毎のアンケートを行うことで評価した。

アンケート調査においては、各情報サービスでそれぞれの技術を説明しながら、実際に操作をするなど体験してもらい、各技術の目

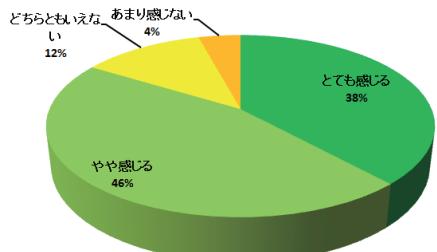


図2 『「変身カメラ」ではプライバシが保護されていると感じるか?』

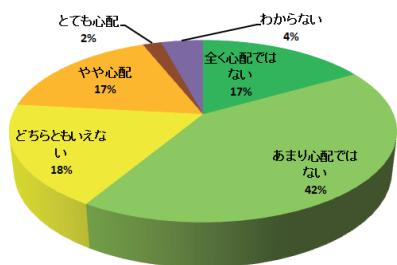


図3 『「変身カメラ」の技術が社会的に広まることは心配か』

指しているものを理解してもらうようにした。

その後、アンケート用紙に記入してもらうことで、結果を収集した。回答は、各項目について5点法で行った。

### (b) アンケート結果

3回のイベント合わせて、216件の回答を得た。ただし、すべての設問に回答していないものも含まれていた。

『「変身カメラ」について理解できたか?』という設問に対して、92.0%が「理解した」「まあ理解した」と回答しており、技術内容の説明は適切に行えたと考えられる。

『「変身カメラ」ではプライバシが保護されていると感じるか?』という設問には、83.4%から肯定的な回答が得られた(図2)。また、『「変身カメラ」の技術が社会的に広まることは心配か』という設問には、58.4%が心配でないと回答した(図3)。さらに、『防犯カメラと比べて「変身カメラ」はプライバシ保護を重視しているか』という設問には、79.2%が肯定的な回答をした。これらのこ

とから、本研究課題のプライバシ保護技術は、実際に技術を体験した一般の来場者からも、一定の理解が得られており、社会的に受容されうることが分かった。

情報サービス毎に『この技術を便利と感じるか』『この技術はプライバシ上心配か』を尋ねた設問では、「にぎわいマップ」では74.2%と67.7%、「デジタルラジオラマ」では76.7%と67.3%、「シースルービジョン」では74.1%と55.2%がそれぞれ肯定的な回答を行った。「シースルービジョン」で、『プライバシ上心配か』に対する肯定的な回答が少なくなっているが、これは、「シースルービジョン」が物陰の情報も見えるようにする技術であるため、隠れていても見られると感じたことが原因と考えられる。実際には、設置したカメラから可視である公共領域内の映像を提示しているが、来場者には多少抵抗感があったと考えられる。その他については、3/4程度が有用と評価しており、また、2/3程度の来場者がプライバシ上の問題を感じていないことから、各情報サービスで示した技術にも、一定の有効性およびサービスの受容性があると言える。

### c. 今後の展望

当初計画した研究開発を行い、実証実験を通じてその社会的な有効性、受容性を調査した。今回の実証実験では、実施施設である新風館が、マイクの設置により一般の来場者が悪印象を抱くことを懸念したため、音声に対するプライバシ保護技術の実験が行えなかつた。この点については、技術的な観点のみでは解決できない。プライバシ保護技術の安全性についてより理解してもらえるよう広報活動を続けることが必要である。

プライバシを扱うためには、各個人が情報提供に同意することも必要である。公共空間において、個々に同意を得ることは、困難かつ利用者にも不便であることから、新たなプ

ライバシ同意方式を考案する必要がある。例えば、個人が情報を収集される環境にあることを知りながら、あえてその環境下に自主的に入るのであれば、その時点での情報提供に同意したと理解できる「環境オプトイン方式」の普及など、関連する法制度も含めた整備が今後の課題である。

## 2. プライバシ情報の管理

### a. 概要

カメラから得られる映像やマイクから得られる音声などのセンサ情報には、被写体の個人情報が含まれており、そのまま公開することはプライバシ保護上問題がある。そこで、画像情報、音情報のそれぞれについて、パターン認識技術を適用し、プライバシ情報を分離する手法を開発した。

### b. 研究目的

USN のセンサノードは被写体の個人情報を被写体の同意を得ずに無断で観測できるため、センサ情報の利活用にはプライバシの問題が生じる。センシング Web は、USN の各センサから得られるセンサ情報を Web と同様の利用形態の下に利活用するための枠組みであり、プライバシの問題はセンサ情報の流通を妨げる原因となる。そこで本研究では、センサからデータを出力する段階でプライバシ保護処理を行うことで、この問題に対処する。

アプリケーションにおいては個人情報が必要なことが多いので、センサ情報を処理して、プライバシ情報を含まないシンボル情報抽出する手法の実現を目指す。

このようなセンサとプライバシ保護処理を一体化した技術を「安心センサ」と名づけ、カメラ画像に対しては「変身カメラ」、マイク音声に対しては「変声マイク」として技術開発を行う。

### c. 研究成果

#### (1) 画像情報に対するプライバシ情報管理

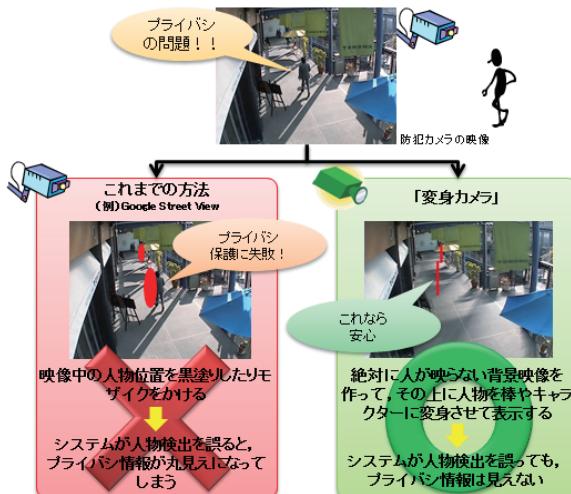


図 4 「変身カメラ」のコンセプト

#### (a) 変身カメラの提案

カメラは人物の姿をそのまま映像データとして出力するため、このデータをそのまま公開・共有すると、被撮影者のプライバシ問題を生じる。したがって、映っている人物が誰であるかが分からないように処理を施した映像や、人の有無や人数などといったシンボルデータや数値データに変換して公開しなければならない。このようなプライバシ保護された映像やデータを生成するためには、画像処理・パターン認識技術が利用できる。例えば、プライバシ保護映像を作成する際には、元の映像に対して人物検出処理を施し、検出された位置にモザイクや黒塗りを行う方法が一般的に用いられる。しかし、このようなアプローチの場合、図4に示すように、人物検出に失敗した際に人物の姿が隠蔽されず、プライバシの侵害が起こる。この人物検出処理は、画像処理やパターン認識の分野で継続的に研究されており、本プロジェクトでも研究を進めているが、任意の映像に対して100%正確に動作する手法を実現するのは不可能である。そのため、生成した映像がプライバシ保護されていることを保証することはできない。

そこで本プロジェクトでは、カメラが一般に固定設置されており、その映像を長期間撮り続けることは容易であることに着目した映

像生成手法を提案した。この手法では、まず長時間観測によって蓄積した画像をもとにその瞬間の背景画像を生成する。一方で、元の画像に対して人物検出処理を適用する。そして、この背景画像上で、人物が検出された位置にキャラクタを重畠描画してプライバシ保護画像を生成する。この手法により、人物検出処理に失敗が生じた際にも、キャラクタの描画位置のずれや描画が行われないなどといったことは起こるが、映像中に人物の姿が見えてしまうことはなく、プライバシの保護を保証することができる。

本プロジェクトは、この提案手法を実装するとともに、この手法を適用したカメラを「変身カメラ」と名付け、積極的に情報発信するとともに、多数の方と意見交換を行った。

#### (b) 長時間性に着目した要素技術開発

(a)で述べた手法によって、人物検出処理の性能が不十分であっても、プライバシを保護することが可能となった。しかし、人物の位置や人数などの正確性を高めるためには、このような要素技術に関する研究も並行して進める必要がある。本プロジェクトは、人物検出処理と背景差分処理について、カメラの長時間映像を解析することによって次第に性能が向上する処理についても研究を行った。

#### (2) 音情報に対するプライバシ情報管理

##### (a) 遠隔発話音声認識システム

センサとしてのマイクロホンは接話マイクではなく、発話者から離れた位置にある。そのため、音声に含まれる言語的なプライバシ情報を抽出するためには、高精度な遠隔発話音声認識技術が必要となる。話者位置を推定することで、高精度な遠隔発話音声認識が可能になるので、話者位置推定技術を開発し、その結果に基づいて、遠隔発話音声認識の高精度化を行った。マイクロホンアレイネットワークとニューラルネットワークによる識別器を用いて、話者方向推定を行うことで、平

面上で推定誤差 20cm（3 次元空間では 23cm）の話者位置推定可能技術を開発した。

#### (b) 声質変換・音声除去システム

「変声マイク」の一形態として、収録した音声に含まれる個人性情報からプライバシ侵害が発生することを避けるために、音声信号内の個人特性を分離して他者の特性に置き換える声質変換システムを開発した。また、周囲環境音をセンサ情報として活用するために、収録した音声に含まれている発話音声を除去する音声除去システムを開発した。具体的には、既存の大規模音声データベース、雑音データベースを使用して、ガウス混合モデル(GMM)に基づく統計的な手法を開発した。

#### (c) 音声認識結果からのプライバシ情報抽出・削除

「変声マイク」の別の形態として、音声認識結果に対してプライバシ情報を保護するために、音声認識結果から話者固有の情報を抽出し、それを削除または置き換える技術を開発した。まず、大規模テキストに対して「半教師あり学習」（正解ラベルが一部にしか付与されていない学習データによる統計モデル学習法）により固有名詞（人名）を自動的に抽出する。そして、抽出された人名に対する処理として三つのレベルを想定し、利用者のアクセスレベルに応じてプライバシ情報を提供するシステムを構築した。

#### d. 今後の展望

画像情報、音情報からプライバシに関連する情報を除去する手法を開発した。現在、環境側に固定設置したセンサを想定しているが、今後はモバイル端末等が普及することから、可動型のセンサから得られる情報のプライバシ保護が課題として挙げられる。また、それぞれのセンサ情報において、個人の情報を除去しても、他の情報と組み合わせることでプライバシ情報が漏れる場合がある。例えば、

個人の家の前の映像では、誰が出入りしたかが分からなくても、その家の家族構成等から、個人が推定できてしまう場合がある。このように、より広範な情報連携におけるプライバシ保護の扱いも、今後の課題である。

### 3. センサ情報の共有

#### a. 概要

センサにより自動収集される情報を公開して利活用するために、サービス仕様、情報要求仕様、サービスと情報要求のマッチング規則を記述する枠組みの作成と、実際に情報提供、利用を行うシステムの開発を行った。

#### b. 研究目的

センシング Web は、センサにより自動収集される情報を源泉として、それを不特定多数のユーザがオープンに活用できるようにする新形態のシステムである。このようなシステムでは、既存の Web システムと比べて扱う情報の階層構成や構造が異なることを踏また上で、センサ情報の利活用におけるサービス需給をマッチさせる技術が必要である。このため本研究では次の 3 つの基本技術を確立する。

- (1) センシングノード群の提供するサービスの仕様記述
- (2) センシング Web に対する情報要求の仕様記述
- (3) サービス仕様と要求仕様をマッチングさせる情報変換・伝達規則

さらに本研究では、これらの形式でセンサ情報を記述し、情報を公開、利用するプロトタイプシステムを構築する。応用例としては、カメラで撮像した映像を原情報とするものと、物体の有無等の物理量を原情報とするものの両方を対象とする。

#### c. 研究成果

- (1) センサノード群が提供するサービスの仕様記述  
種類の異なるセンサに対しても、得られる

データの差異に依らず、統一的にセンサ情報を記述する枠組みを開発した。仕様記述に関しては、ロボット制御に用いられる時空間情報管理技術を調査し、これを環境センササービスの仕様記述と組み合わせた。これによって、統合化サービスとして時空間的な広がりをもつサービス仕様記述を可能とした。例えば、任意の空間に存在する複数の環境センサを検索し、同種のデータをひとまとめにして取得することが可能となる。また、計測間隔等の環境センサの時間的特性を利用して、複数のセンサ情報を同期させることも可能となる。この時、時空間情報に含まれる誤差についての記述を可能としたことで、統合化サービスに関する品質仕様が記述可能となったと言える。ロボット制御用の技術はオブジェクト指向で設計されていたため、マークアップ言語 XML のスキーマ（語彙集合の設計書）へのマッピングを行った。

### (2) センシング Web に対する情報要求の仕様記述

空間的な広がりをもつ環境センサノード群に対して、環境情報を利用者が要求するための適切な語彙集合を開発した。任意の空間を指定する語彙（例えば、球形状を表す語彙<globular>と座標を表す語彙<coordinate>を組み合わせることで球空間を指定することができる）を用いることで、その中に存在する複数地点の環境データを一覧で取得することができるようになる。また、一覧で取得した環境データを組み合わせることで得られる情報を要求するために、上位の抽象度をもった語彙が利用できる。語彙集合の設計には、マークアップ言語 XML のスキーマ技術を用いた。

### (3) サービス仕様と要求仕様をマッチングさせる情報変換・伝達規則

複数の環境センサノードが任意の空間内で連携することを前提に、(1)で開発したサービス仕様記述と(2)で開発した要求仕様記述の

間の時空間的な対応規則を変換言語（XSLT）で規定した。空間情報に関しては、サービス側と要求側で異なる座標系を用いた場合にも自動変換できるようにした。同様に時間情報においては用いられる標準時の自動変換を可能とした。要求に対してサービス側の情報が不足した場合には、空間的に近くにある別のサービスを検索し、これを用いて補間することにした。また、保守性を高める観点から、座標系や標準時の自動変換を行うモジュールと、時空間統合（別サービスによる情報補間や環境データ間の同期）を行うモジュールとに分離して開発した。

### (4) センサ情報提供システムの構築

開発した3つの基本技術を実装し、実際に、センサから得られる情報を公開する仕組みを構築した。HTTP プロトコルを用いた状態を持たない問い合わせ形式（REST）により外部からの要求に応える Web サーバを構築した。また、この情報提供システムを使って、センサ情報を取得して動作するサンプルアプリケーションを作成した。

### d. 今後の展望

センシング Web のセンサ情報を公開し、利活用するための技術を開発した。センサ情報を公開するシステムを構築し、実際にデータ公開を行っているが、現状では、プロジェクト外からのアプリケーション作成などの貢献事例を得るに至っていない。これは、データ公開の時期が遅くなったり、およびデータ公開の広報が不十分であったことが原因である。今後、本研究の技術を広く公表し、多くの人に利用してもらうことで、有効性の検証を行うことが課題である。

## 4. 観測型コンテンツの提示

### a. 概要

センシング Web では、実世界に設置された多数のセンサから得られる情報の利用を想定

している。このような多数のセンサ情報をそのまま利用者に提示しても、個々のデータの意味や、データ間の関係の把握が困難である。そこで、本研究では、提供されるセンサ情報をコンテンツとして効果的に提示する技術を開発した。

#### b. 研究目的

センシング Web で開発した技術により、世界中に分散する様々なセンサ情報に対し、プライバシ情報がフィルタリングされた状況で、同一のプロトコルを用いてアクセスすることが可能となる。センサ情報の一つの特徴は、実世界を観測した多種類で多量のデータが実時間で生成され続ける点である。また、センサ情報は通常、そのセンサが設置された空間、および観測された時間と結びついたデータとなっている。このような時空間的な特性を持った膨大なセンサ情報を利用者が利活用できるには、その特性を反映したコンテンツとして、利用者に対して効果的にデータを提示できなければならない。このようなコンテンツのことを観測型コンテンツと呼ぶ。

本研究では、実世界に分散配置された様々なセンサから得られるセンサ情報と、地理情報等の広域環境情報を有機的に統合利用すると共に、これに基づいて利用者からの要求に応じてコンテンツを動的・適時的に提示する情報サービスを構築することを目的とする。

#### c. 研究成果

観測型コンテンツを提示する情報サービスとして、以下の 3 種類を提案・実装した。

##### (1) にぎわいマップ

センシング Web のセンサ情報を統計情報化して、2 次元の地図上に提示する情報サービスである「にぎわいマップ」を構築した。この情報サービスでは、2 次元の地図を一定の大きさのブロック単位で分割し、それぞれのブロック内で、過去数秒の間に観測された人

物の数を累計し、地図上の各部分がどの程度混雑しているか（にぎわっているか）を提示する。センシング Web で開発したセンサ情報提供の仕組みを利用して、各ブロックに含まれるセンサ情報を取得し、一定期間累計した結果を、Web ブラウザで閲覧可能な形式に変換して提示するよう実装した（図 5）。

個々のセンサ情報は、観測された細かな位置情報をを持つが、これをブロック単位に累計（統計情報化）することで、おおまかな人物分布として要約提示することができる。ブロックの大きさは、スライダバーを用いて利用者が変更可能となっており、要約の粒度調整が可能である。また、利用しているセンサ情報は、プライバシ保護処理がなされたものであるが、統計情報化することで、個々の人物情報を識別することができなくなるため、より個人性の少ない情報提示方法となっている。

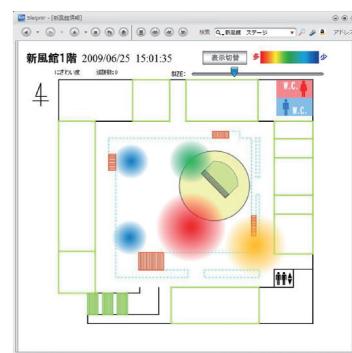


図 5 「にぎわいマップ」

##### (2) デジタルジオラマ

実空間を反映した 3 次元モデルを作成し、その実空間に設置されたそれぞれのセンサから得られる情報を 3 次元モデル上に配置して提示する「デジタルジオラマ」（図 6）を開発した。利用者は、3 次元モデル内で、任意の位置から任意の方向を見ることができる。また、3 次元モデルでは、事前に撮影した画像をテクスチャとして用いているが、デジタルジオラマ内で、カメラセンサ付近に移動すると、カメラセンサから得られる現在の背景画像をデジタルジオラマ上に重畳して見るこ

とができる。さらに、カメラの視野をデジタルジオラマ上に表示する機能も実装した。時空間的特性をもつセンサ情報を、一つの3次元空間に配置し、任意視点から提示することで、利用者は実世界の実時間の情報を、空間情報として容易に把握することができる。

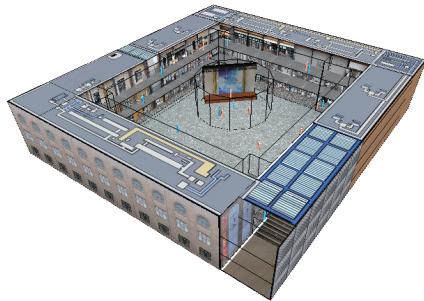


図 6 「デジタルジオラマ」

デジタルジオラマには、プライバシ制御の機能も実装した。人物等の移動物体に対しては、プライバシ保護処理が行われている「変身カメラ」を利用しており、人物画像の代わりに個人情報を含まないキャラクタとして表示される。一方で、例えば同行者とはぐれて相手を探したい場合では、相手を特定する必要がある。これに対しては、あらかじめ同行者グループが同じRFIDタグをもつことで、同行者に対してのみ、個人を識別可能な形で情報提示する仕組みを実装した。このようなプライバシ制御の考え方の導入により、利用者に応じた柔軟な情報提供が可能になる(図7)。



図 7 「デジタルジオラマ」におけるプライバシ制御

### (3) シースルービジョン

環境に設置したセンサの情報をを利用して、利用者からは物陰となっている場所の情報を提示する「シースルービジョン」を開発した。利用者とのインターフェースには、カメラ付きのモバイル端末を用いた。利用者の持つ端末の位置姿勢を計測し、端末のカメラ映像に対して適切な位置、大きさで、環境側のカメラの映像を重畠することで、情報付加した映像を提示した。モバイル端末と固定センサの情報を連携し、複合現実感技術を用いることで、実世界での利用者視点に応じた情報提示を可能とした(図8)。プライバシ保護に関しては、「変身カメラ」を利用して人物のシンボル化を行った。物陰の情報が見えることから、実証実験のアンケートでは、プライバシ上の懸念が見られたが、利用しているセンサ情報は変身カメラと同様である。利用者視点で提示していることで現実感が増し、逆にプライバシ上の課題が生じていると考えられる。

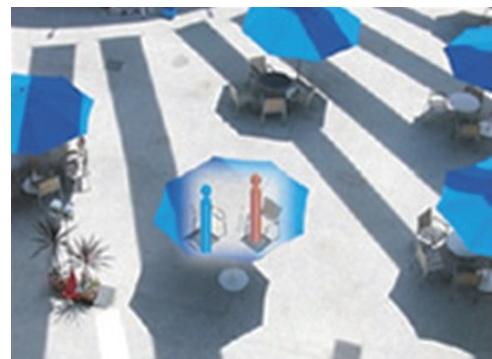


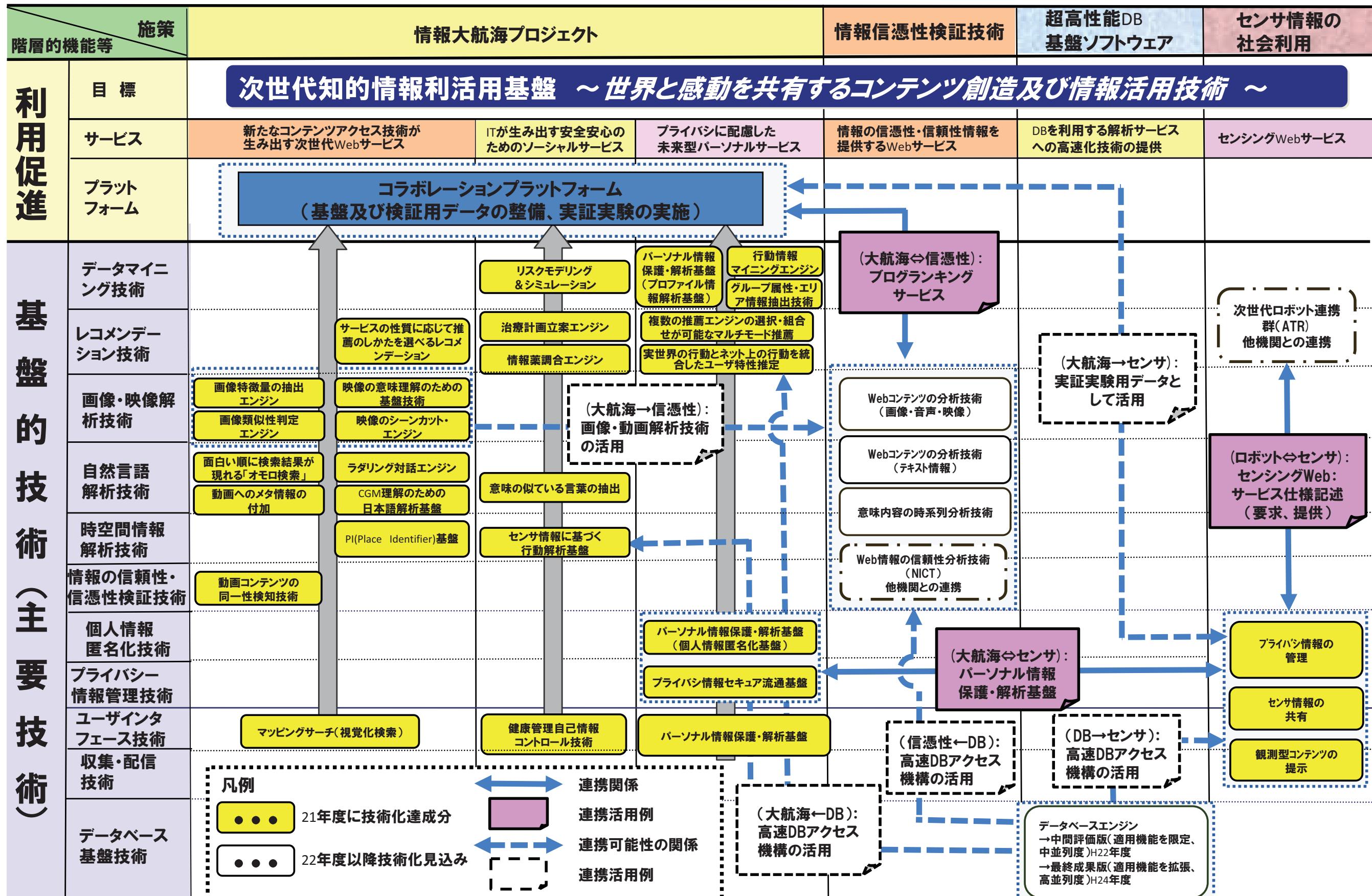
図 8 「シースルービジョン」

### d. 今後の展望

本研究では、多数のセンサ情報を効果的に提示する技術を開発した。プライバシ保護技術により、センサ情報から個人の情報を除去してコンテンツとして利用していたが、個人を特定し、個人適応したオーダーメードサービスが有用となる場合も考えられる。現在は、RFIDタグを用いたグループ単位でのプライバシ制御を実現しているが、匿名情報と個人情報をより細かく切り替えることが可能な仕組みについても、今後検討する必要がある。

## 4. 連携施策群の基盤技術

### 4.1 情報の巨大集積化と利活用基盤技術開発（技術連携マップ）



## 4. 連携施策群の基盤技術

- 4.1 情報の巨大集積化と利活用基盤技術開発（技術連携マップ）
- 4.2 技術連携マップの基盤技術説明
  - 4.2.1 情報大航海プロジェクト
  - 4.2.2 情報信憑性検証技術
  - 4.2.3 超高性能データベース基盤ソフトウェア
  - 4.2.4 センシング Web
- 4.3 基盤技術の活用例
  - 4.3.1 行動情報に基づいた最適な情報やサービスの提供
  - 4.3.2 高度な予防型安全管理システムの実現
  - 4.3.3 生活習慣に関する情報等に基づいた「情報薬」の提供
  - 4.3.4 ブログランキングサービス
  - 4.3.5 センシング Web

## 4.2 技術連携マップの基盤技術説明

### 4.2.1 情報大航海プロジェクト

画像・映像解析技術

次世代Webサービス

#### 画像特微量の抽出エンジン

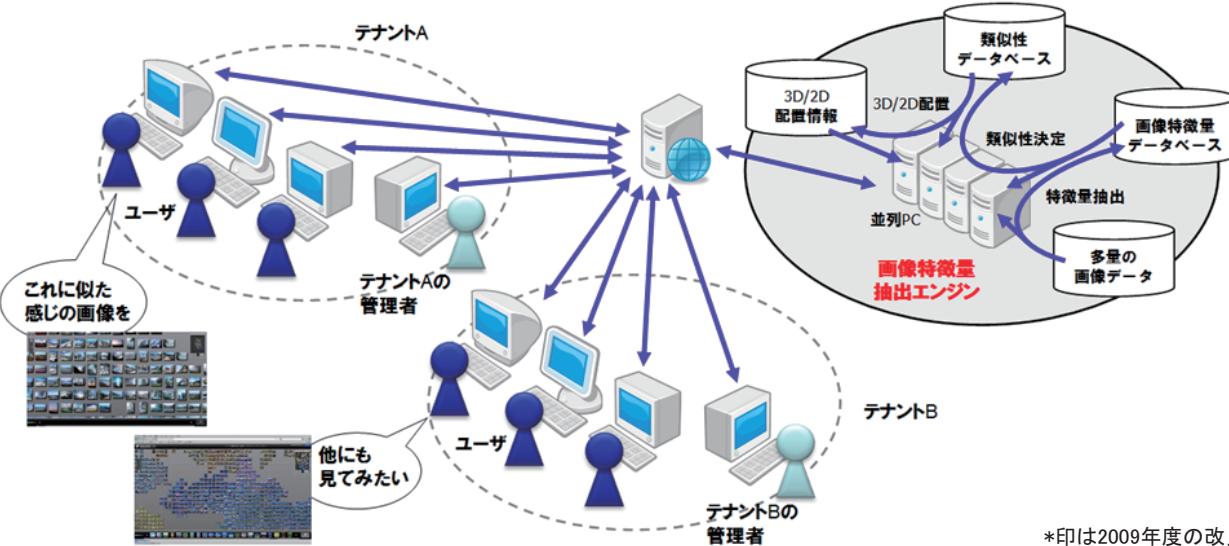
膨大な画像群の、特微量抽出、類似性計算、及び3D/2D配置情報生成を高速実行するエンジン

##### 技術の概要

- 画像を検索する時、「キーワードを用いたくない」や「多数の候補を見ながら選びたい」といった要求が多くあり、大量の画像群が表示された空間をユーザーが散歩しているような画像アクセスが求められている
- そのために、登録された大量の画像から抽出された色、輪郭線、配色等の特微量を用いて画像間の類似性を計算し、類似性データベースを作成した上で、3D/2D空間への画像群の配置情報を生成する
- 特微量抽出と画像間の類似性計算を高速化するために、並列処理技術及びハッシュ技術を採用している
- マルチテナント化、導入運用の自動化、追加特微量組込み機構の実現により、簡便に幅広いサービスへの導入が可能である

##### 技術の特長

- 画像特微量の抽出
  - 人の感覚に沿って選ばれた、色、輪郭線、配色等の特微量を抽出する
- 高速な解析処理
  - 画像の特微量抽出と画像間の類似性計算を高速化するため、並列処理技術及びハッシュ技術を採用しており、画像10万枚の処理(画像特微量データベース及び類似性データベースの作成、3D/2D空間への配置情報の生成)が28時間で可能であり(1CPUの場合)、画像を頻繁に追加・削除するサービスへの適用も可能である
- 画像間の類似性を活用したユーザーインターフェース
  - 多様なユーザー要求に対応するため、画像群を表示する時に、あらかじめ決められた方式で生成された配置だけでなく、ユーザーが指示した基準での再配置も可能である
  - 同一対象物を撮影した複数画像等のグルーピング機能、属性情報の絞込み機能、ズームアウト機能、配置アルゴリズムの改善、サムネイル表示のための処理高速化等の性能改善等を実現している
- マルチテナント化、導入運用の自動化、追加特微量組込み機構の実現\*
  - 本エンジンの動作に必要なサーバ群をマルチテナント化し、導入企業の利用コストの低減を実現している
  - 導入及び運用の自動化を実現し、機動的な運用を可能としている
  - 必要とされる特微量を追加できる機構の実現により、幅広いサービスへの適用を可能としている



\*印は2009年度の改良・追加技術

## 4.2.1 情報大航海プロジェクト

画像・映像解析技術

次世代Webサービス

### 画像類似性判定エンジン

#### ユーザビリティの高いクエリ画像による順位付き類似画像検索

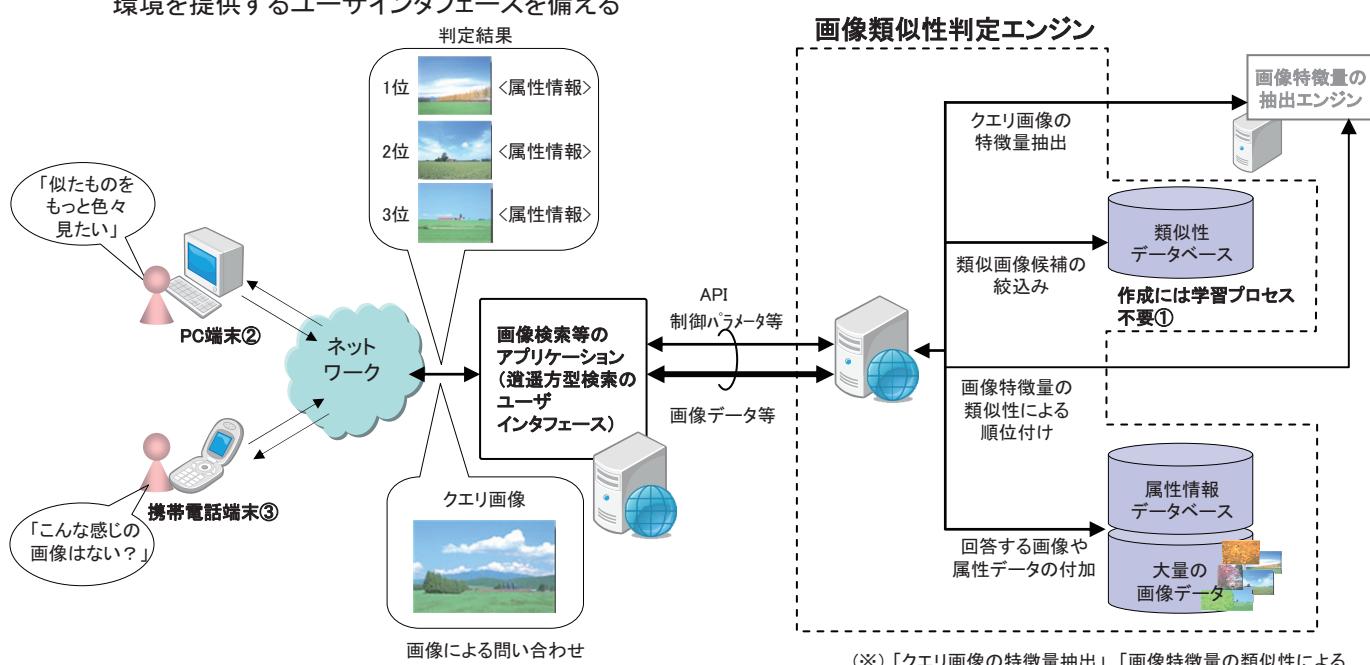
##### 技術の概要

- 登録済みの大量の画像群の中から、ユーザーによるクエリ画像(検索用に指定された画像)に類似する画像の探索を可能とする
- クエリ画像から画像の特徴量を抽出し、登録済みの大量の画像群<sup>(注)</sup>の中から、類似性に基づいて検索することにより、順位付けられた類似画像とそれらの属性情報を出力する
- 上記技術を活用し、PCと携帯電話端末のいずれからも、クエリ画像に類似した画像を検索できるデモ用アプリケーションを実現した

(注) 検索対象となる全ての画像の特徴量をあらかじめ抽出し、類似性データベースに蓄積しておくことにより、検索に備える

##### 技術の特長

- 学習プロセスが不要
  - ハッシュ技術による画像の類似性判定法を利用し、時間と手間の掛かる学習プロセスを排除したので、機動的に応用できる
- 遙遠型検索環境との親和性と利用者の満足度向上
  - 「画像特徴量の抽出エンジン」とデータ共有できるため、遙遠型検索のユーザインターフェースと併用して、多量の候補を提示することにより利用者の満足感を高められる
- 携帯電話端末でのサービス提供
  - PCと携帯電話端末の両方に対応可能であり、携帯電話端末に対してもユーザビリティの高い検索環境を提供するユーザインターフェースを備える



## 4.2.1 情報大航海プロジェクト

自然言語解析技術

次世代Webサービス

### 面白い順に検索結果が現れる「オモロ検索」

インターネット上の主観的な「面白さ」をよりよく反映した検索結果の提示技術

#### 技術の概要

- ユーザーの興味／関心の度合い<sup>(注)</sup>を「オモロ度」と定義して、Webクロールによって収集したテキスト情報等からコンテンツの「オモロ度」を算出する
- 算出された「オモロ度」に基づき、「オモロ度」の高いコンテンツ、すなわちユーザーの興味／関心を強く引いているコンテンツが上位表示される検索エンジンであり、多くの人に必要とされるコンテンツが上位表示される一般的な検索エンジンとは一線を画す
- ユーザーは「オモロ検索」により、少ない操作で「オモロ度」の高いコンテンツに素早く到達することができる

(注) 興味／関心の判断基準については、共通技術「動画へのメタ情報の付加」で収集したメタデータ等を使用(ただしアルゴリズムは非公開)

#### 技術の特長

##### 評価情報の反映機能

…①

- コンテンツ情報(動画共有サイトから収集できる情報など)だけでなく、ユーザーによる評価情報も、Webクロールによって収集し、メタ情報とする
- 新たに登録されたコンテンツを直ぐに検索対象とできるよう、インデックス化の処理を高速化する

##### 正規化機能

…②

- 正規化辞書を構築し、表記揺れ／新語に対応する

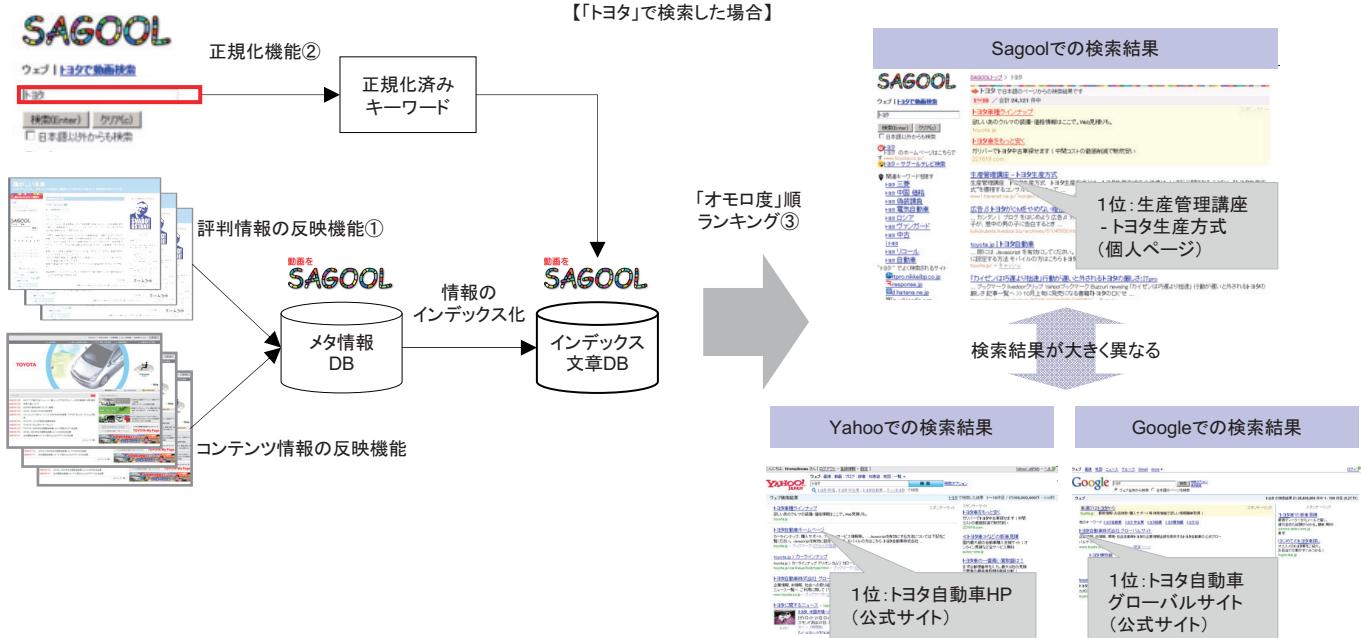
##### 「オモロ度」順ランキング

…③

- ユーザーの興味／関心の度合いに基づき「オモロ度」を算出し、「オモロ度」の高いコンテンツを上位表示する

##### 重要語の自動抽出機能\*

- キーワードの入力補助、スニペットの補完などに利用できる重要語を、検索対象の文章を基に自動抽出する。



\*印は2009年度の改良・追加技術

igvp-technology@meti.go.jp

## 4.2.1 情報大航海プロジェクト

自然言語解析技術

次世代Webサービス

### 動画へのメタ情報の付加

メタ情報の生成により動画検索のヒット率を向上させる

#### 技術の概要

- Webクロールによって、複数の動画共有サイトから動画情報を収集し、加えて、動画に関する評判情報を、動画共有サイト外からも収集する
- 動画共有サイトから収集した動画情報と、動画共有サイト外から収集した動画に関する評判情報を、「メタ情報」として動画に付加することで、様々な言葉で動画の検索が可能になると同時に、有効なレコメンド(テキストマッチ)が可能となる

#### 技術の特長

##### ● 複数動画共有サイトへの対応

…①

- YouTubeやAmebaVisionをはじめとした、90以上の国内外の動画共有サイトに対応し、動画情報を統一的に扱えるデータベースを構築する
- 動画データの増加傾向を鑑み、1億件以上の動画に対応すべく、データベースの大規模化を行う\*

##### ● 評判情報の収集機能

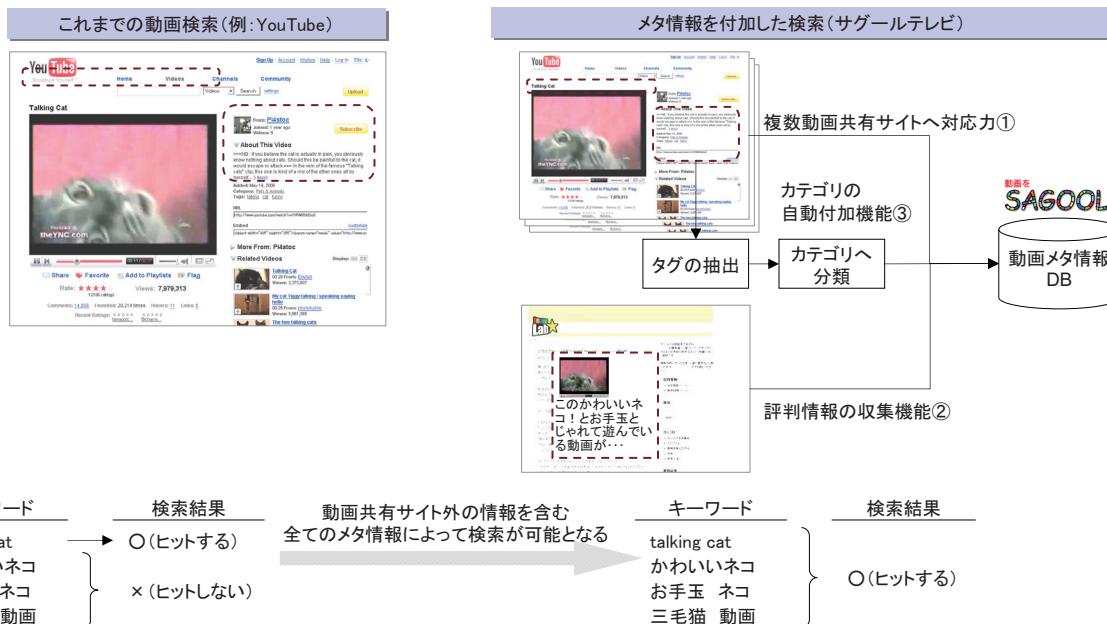
…②

- 動画共有サイト内の情報(動画投稿者が付加)のみならず、Webクロールによって、ブログなどに掲載されている動画の評判情報(動画視聴者が作成したテキスト情報)を自動判別し、収集・付加する

##### ● カテゴリの自動付加機能

…③

- 動画投稿者があらかじめ付加しているタグを参照することで、動画に対して、複数の動画共有サイトで異なるカテゴリを統一的な15種のカテゴリに自動分類する\*



\*印は2008年度の改良・追加技術

## 4.2.1 情報大航海プロジェクト

情報の信頼性・信憑性検証  
技術

次世代Webサービス

### 動画コンテンツの同一性検知技術

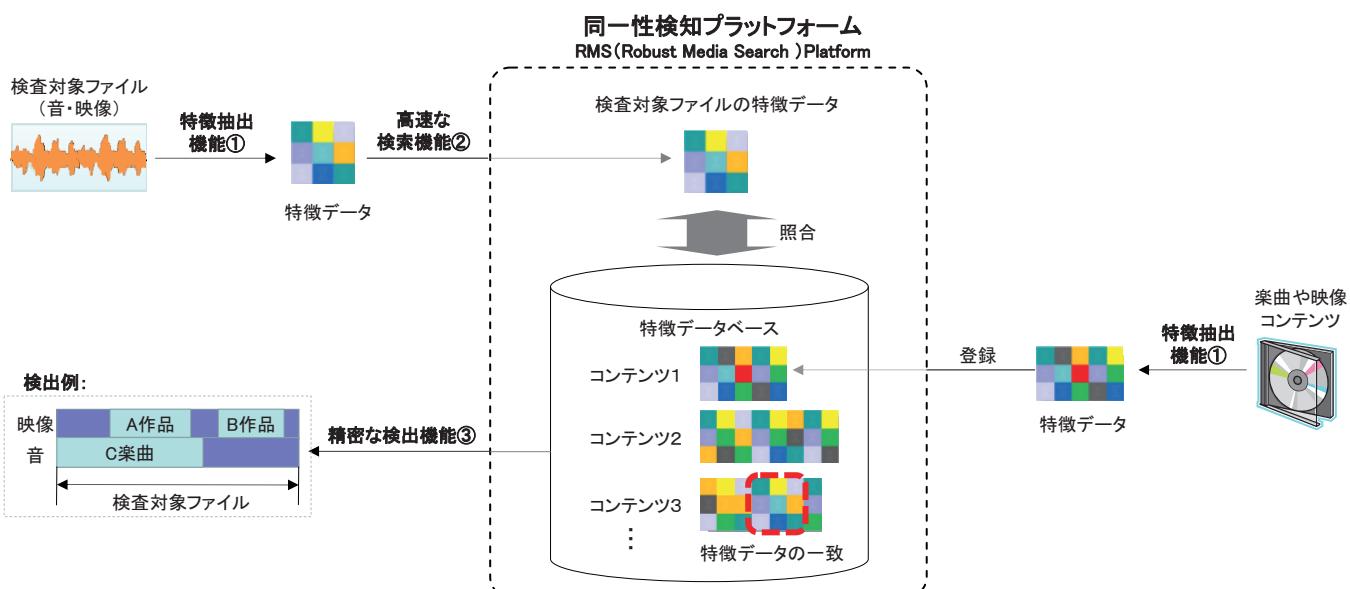
#### 既知の音や映像の使用を特定する技術

##### 技術の概要

- 大量の音や映像ファイルを対象として、既知のコンテンツが存在するか否かを検知する技術である
- 音や映像ファイルから抽出する特徴データ(音の波形値や映像の画素値を基に算出される値)を利用することにより、コンテンツ間の高速な照合と同一性検知が可能となる
- 本技術の利用例としては、ネットワーク環境のプロードバンド化により、大量の動画コンテンツがインターネット上に存在している現在の状況において、各コンテンツが正規に権利処理されたコンテンツか否かを判別する際の利用等が考えられる

##### 技術の特長

- 動画の変化や歪みに強い特徴抽出機能 ...①
  - 検査対象のファイルに様々な編集(別の音の重なり、テロップなどの重なり、音質や画質の劣化、編集や加工など)が加えられている場合でも、有効な同一性検知に必要な特徴データを抽出可能である
- 高速な検索機能 ...②
  - 現在実施中の商用サービスの一例においては、特徴データベースに約2,000時間相当のコンテンツが登録されている状態で、1日に約100万分相当(例:平均4分の映像ファイルの場合、25万件に相当)の処理を行っている
- 精密な検出機能 ...③
  - 音や映像ファイルの先頭からのみでなく、「検査対象ファイルの何秒目から何秒目に、登録コンテンツの何秒目から何秒目が使用されているか」を、音と映像それぞれについて精密に検出する



## 4.2.1 情報大航海プロジェクト

レコメンデーション技術

次世代Webサービス

### サービスの性質に応じて推薦のしかたを選べるレコメンデーション

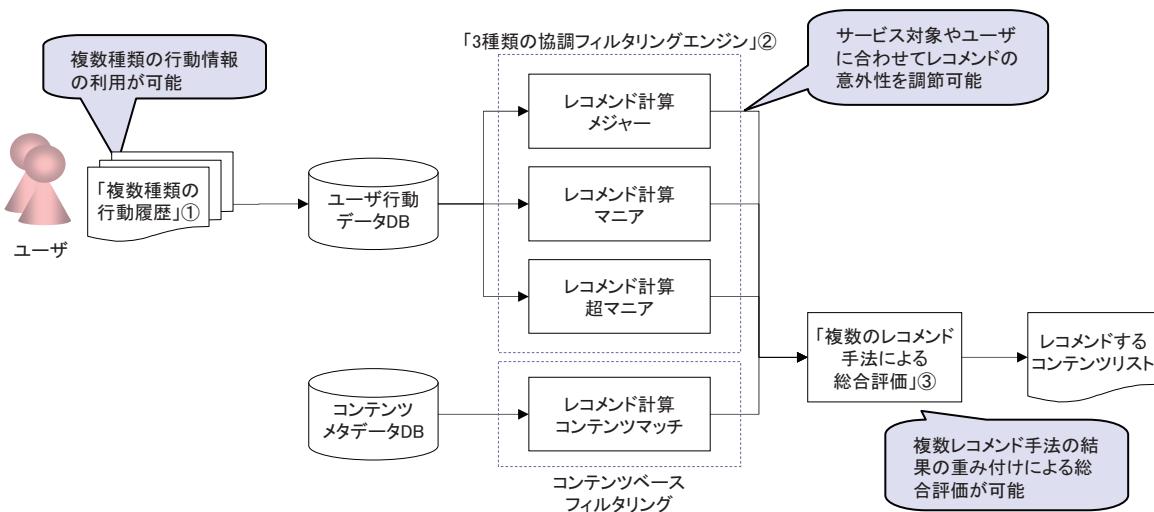
特性の異なる3つのエンジンで、マニアックなもの、意外なものもレコメンド

#### 技術の概要

- 複数のレコメンド技術を利用してすることで、ユーザの行動履歴やインターネット上に存在する様々な情報から、個別ユーザの嗜好と合致した情報を推薦する
- これまで、個別ユーザの嗜好性を反映した情報をレコメンドするためには、一定量の行動履歴の蓄積が必要とされていたが、本技術においては、複数の手法によるレコメンド結果を、行動履歴の蓄積量に応じて重み付けをして総合評価することによって、ユーザのサービス利用初期から個別ユーザの嗜好性を反映したレコメンドが可能となる

#### 技術の特長

- 「複数種類の行動履歴」\*
  - ECサイトでの商品検索、動画再生完了、お気に入りへの追加といった、ウェブサイト上におけるユーザの複数種類の行動情報を蓄積して、ユーザの嗜好性を判断した上でレコメンドを行う
- 「3種類の協調フィルタリングエンジン」
  - メジャー、マニア、超マニアの3種類のエンジンの中から、サービス対象・ユーザ対象に対して最適な計算を事業者が選択することができる
- 「複数のレコメンド手法による総合評価」\*
  - 協調フィルタリングによるレコメンド結果と、コンテンツマッチ（類似しているコンテンツの照合）を行うコンテンツベースフィルタリングによるレコメンド結果を総合的に評価することで、ユーザの行動履歴の蓄積数が少ない場合でも、ユーザの嗜好性を反映したレコメンドを行うことが可能である



\*印は2008年度の改良・追加技術

## 4.2.1 情報大航海プロジェクト

画像・映像解析技術

次世代Webサービス

### 映像の意味理解のための基盤技術

個人作成の静止画像や映像を、効果的に解析・検索するための基盤技術を提供

#### 技術の概要

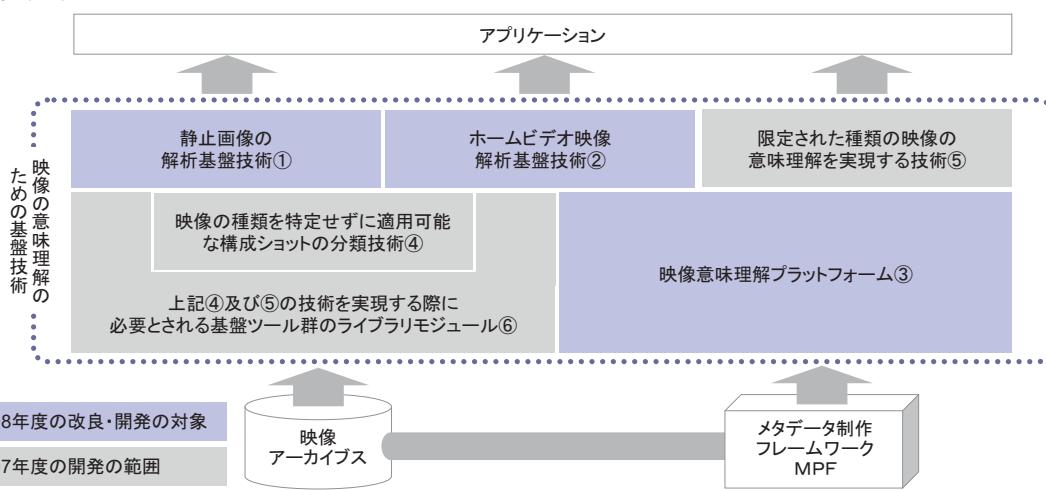
- 一般ユーザがデジタルカメラで撮影した静止画像を、適切な概念カテゴリ(人物、屋外、食べ物など)に分類し、概念カテゴリに関するメタデータを自動付与する
- また、ホームビデオ映像をショットに分割し、各ショットのカメラモーション推定及び概念カテゴリへの自動分類を行い、カメラモーションや概念カテゴリに関するメタデータを自動付与する
- 静止画像や映像などのコンテンツに対して、意味内容を示すメタデータを付与するための共通プラットフォーム(MPF<sup>注</sup>)の提案、及びその共通プラットフォームにおいてコンテンツの検索／表示をすること可能にする

(注) MPF: Metadata Production Framework(メタデータ制作フレームワーク)

#### 全体構成

- 本技術は、2007年度開発「映像の意味理解のための基盤技術」を発展させ、適用対象となる映像の種類を拡大させた技術であり、以下の3つの技術要素群から構成されている

- 「静止画像の解析基盤技術」  
➤ 静止画像を概念カテゴリへ自動分類し、概念カテゴリに関するメタデータを自動付与する  
➤ 概念カテゴリの追加／変更などを容易に行うことを実現するためのツールを開発する\*  
➤ メタデータが付与された静止画像を、複数のモードで選択／表示させるツールを開発する\*
- 「ホームビデオ映像解析基盤技術」  
➤ ホームビデオ映像をショットに分割し、代表フレームを抽出するとともに、カメラモーションを推定する\*  
➤ ショット内の代表フレームを概念カテゴリへ自動分類し、概念カテゴリに関するメタデータを自動付与する  
➤ 概念カテゴリの追加／変更などを容易に行うことを実現するためのツールを開発する\*  
➤ メタデータが付与された代表フレームを、複数のモードで選択／表示させるツールを開発する\*
- 「映像意味理解プラットフォーム」\*  
➤ ホームビデオ映像解析基盤技術、及び情報大航海プロジェクトで開発された他の映像処理技術との連携利用を促進する映像意味理解プラットフォームを提案する  
➤ 上記、映像意味理解プラットフォームに準拠し、映像の検索・再生をWebブラウザ上で行えるツールを開発する



①から③は、2008年度の改良・開発の対象

④から⑥は、2007年度の開発の範囲

\*印は2008年度の改良・追加技術

igvp-technology@meti.go.jp

## 4.2.1 情報大航海プロジェクト

画像・映像解析技術

次世代Webサービス

### 映像のシーンカット・エンジン

#### 映像のシーンカット、メタデータ付与、特徴的な映像区間抽出のための解析技術

##### 技術の概要

● 本技術は、以下の3種類の技術を備えることで、対象とする映像内容に応じた多様な解析を可能とする

- 映像種別に依存しないシーンカット解析技術
- 編集済みの特定映像(サッカー、野球、自然・紀行番組)を対象とした映像の意味解析技術
- 未編集の映像素材コンテンツを対象とした特徴的な映像区間の解析技術

##### 技術の特長

● 「映像種別に依存しないシーンカット解析技術\*」

…①

- 画像解析と音響信号解析の結果を利用してカット位置を判別することにより、低い処理負荷で高い精度のカット位置検出を実現している
- 入力できる映像コンテンツのサイズを1時間に拡大している

● 「編集済みの特定映像を対象とした映像の意味解析技術\*」

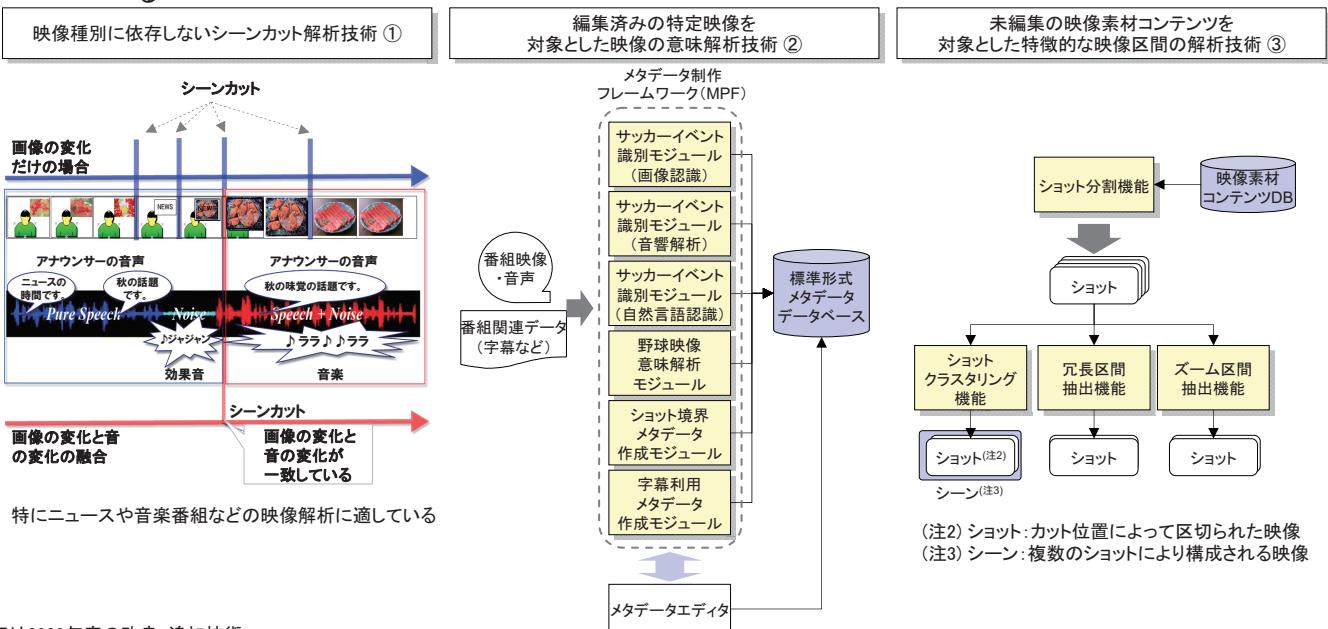
…②

- 映像を内容に基づいて切り分け、その意味内容に応じたメタデータを付与することができる  
(メタデータの例: スポーツのイベント(ホームランなど)、被写体の名前、動作等)
- サッカー、野球、自然番組、紀行番組のそれについて、適合したモジュール<sup>(注1)</sup>を用いる  
(注1) 新たなモジュールをメタデータ制作フレームワーク(MPF)に追加することにより、切り分け・タグ付け可能な映像の種類を増やすことが可能である

● 「未編集の映像素材コンテンツを対象とした特徴的な映像区間の解析技術\*」

…③

- ズーム中の映像区間や、変化の少ない冗長な映像区間を抽出可能である
- 効率的な検索を実現するために、時間情報を元にショット、シーン<sup>(注2)</sup>を構成(ショットクラスタリング)している



\*印は2008年度の改良・追加技術

## 4.2.1 情報大航海プロジェクト

自然言語解析技術

次世代Webサービス

### ラダリング対話エンジン

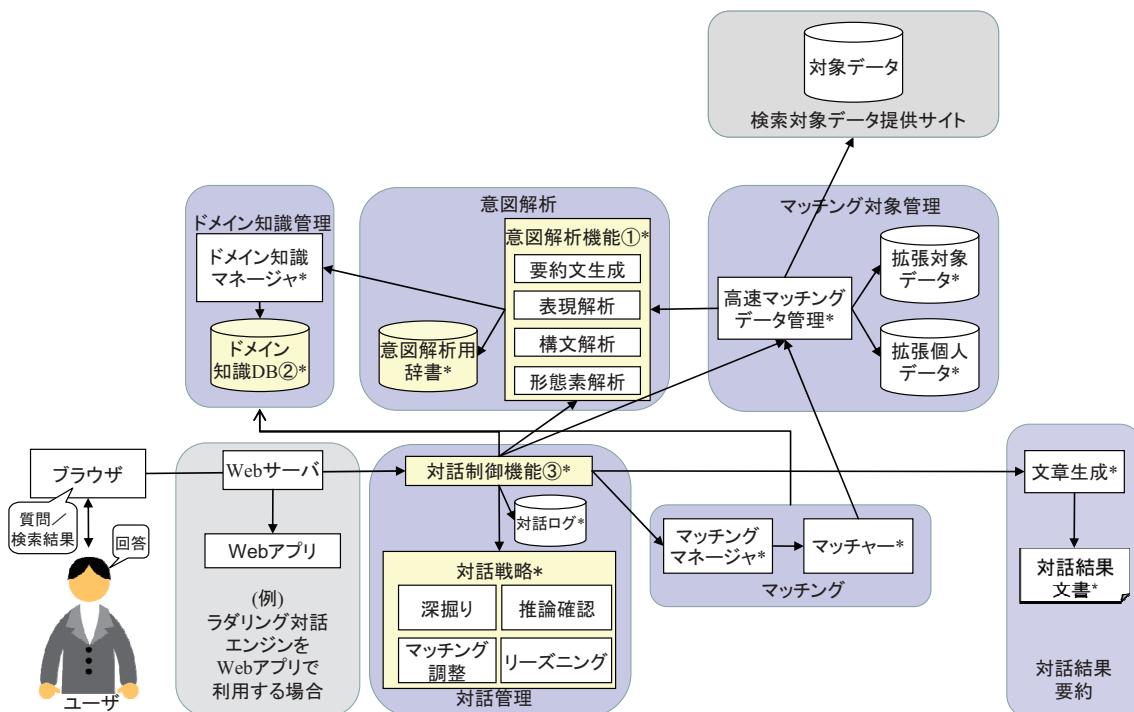
対話により、顧客の「真のニーズ」を“見える化”し、マッチングを行う

#### 技術の概要

- ユーザーの情報検索ニーズが曖昧なときに、ユーザーと対話を進めていくことにより、ユーザーのニーズを顕在化し（「ラダリング」と呼ぶマーケティング調査手法の一種）、適切な情報とのマッチングを実現する
- 対話におけるユーザーの回答文を自然言語解析してユーザーの意図を読み取り、それを基に質問文を生成することで、文脈に沿った自然な対話を実現する

#### 技術の特長

- 意図解析機能\*
  - 意図解析用辞書を用いた構文解析により、ユーザーの回答文から品詞や係り受けだけではなく、意図や感情等も取得し、対話で引用できるように要約文も生成する...①
- ドメイン知識データベース(DB)\*
  - 対話のドメイン(対象分野)に関する知識データベースであり、言葉や言葉同士の関係、特定の言葉に関する質問文等、ドメインに関する全ての情報をオントロジー<sup>(注)</sup>で整理する
  - ドメイン知識を入れ替えることにより、他のドメインにも応用可能である...②
- 対話制御機能\*
  - ユーザーから獲得した情報を基に、複数の「対話戦略(質問文選択のルール)」にしたがって次の質問文を選択する
  - オントロジーを利用してことで、対話中に獲得した情報を考慮した質問文の生成が可能である...③



(注) 情報間の関係を用いて、情報を構造的に表現する方法

\*印は2009年度の改良・追加技術

## 4.2.1 情報大航海プロジェクト

自然言語解析技術

次世代Webサービス

### CGM理解のための日本語解析基盤

本基盤で構築するデータを用いて、日本語解析器のCGM解析精度を向上可能

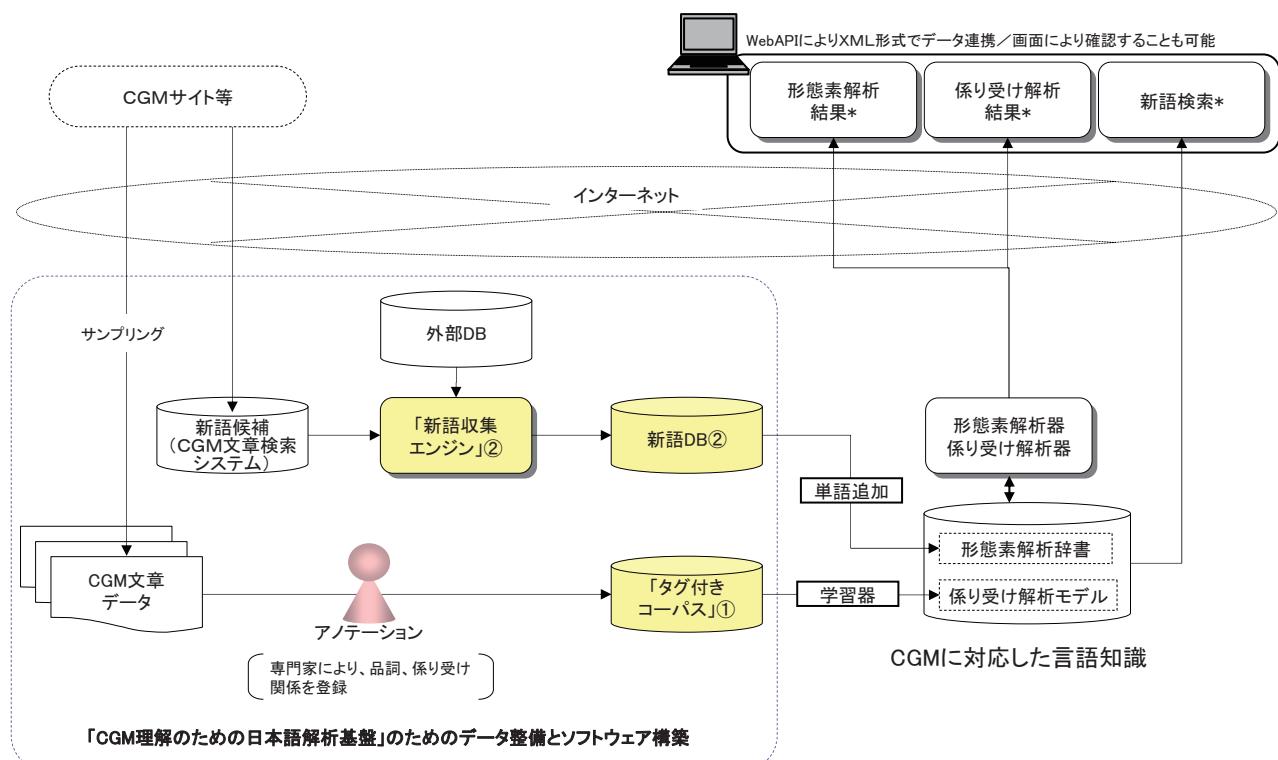
#### 技術の概要

- CGM文章に特有の口語文に対応して「形態素解析」及び「係り受け解析」を精度良く行うため、文法情報（係り受け解析モデル）の学習用データとなる「タグ付きコーパス」（人手で正しい解析を行なった文章DB）を整備する
- CGMにおいて次々と現れる新語をリアルタイムに抽出し、品詞・属性情報を自動で付与する「新語収集エンジン」を開発し、「新語DB」を整備する
- 「タグ付きコーパス」及び「新語DB」を用いて、CGMに対応した言語知識を構築し、これを利用することによりCGMに対して精度が低かった従来の日本語解析器の解析精度を向上させることができる

\*CGM:消費者生成メディア 形態素解析:文を単語に分かち書きし、品詞を付与する処理 係り受け解析:文節間の係り受け関係を解析する処理

#### 技術の特長

- 「学習用データ(タグ付きコーパス)\*」
  - 日本語解析の専門家が約1万文のCGMの口コミ文章を解析し、タグ付きコーパスを整備
  - 2008年度に作成したタグ付きコーパス数は8,500文
- 「新語収集エンジン及び新語DB\*」
  - 新語収集エンジンにより、CGM文章の検索システムで使用された検索語を新語候補とし、新語判定、品詞判定、カテゴリ推定、不適切語判定を自動で行う
  - 2008年に現れた約2,000万語の新語候補から、認知度の高い約50万語を厳選し、新語DBに登録



\*印は2008年度の改良・追加技術

igvp-technology@meti.go.jp

## 4.2.1 情報大航海プロジェクト

時空間情報解析技術

次世代Webサービス

### PI(Place Identifier)基盤

場所に紐付くコンテンツを、より「簡単に」「正確に」提供するためのプラットフォーム

#### 技術の概要

- 同一の場所に関する様々な情報を統合し、異なる地図システム間の情報に関する位置のズレを補正する
- 場所に関する情報は「住所」「施設名」「通称」「位置(座標)」など様々なものがあり、コンテンツによって表現方法も様々であるため、ユーザーが欲しいコンテンツを検索・取得する際の障壁となっていた
- また、カーナビゲーションやインターネットの地図サービスが普及しているが、地図に記載されている道路や建物の形状は地図システムごとに異なることが多いため、ある地図上で登録されたコンテンツを別の地図上で表示させようとすると、位置がずれてしまう問題が生じていた
- 本基盤は、様々な“同じ場所”に関する情報(PI<sup>注1</sup>)を統合する仕組みと、異なる地図システム間で発生する位置のズレを補正する仕組みによって、場所に紐付くコンテンツをユーザーにより「簡単に」「正確に」提供することを可能とする  
(注1) 場所に関する情報(値)をPI(Place Identifier)と呼ぶ

#### 技術の特長

##### PI機能

…①

- 同じ場所を指すPI同士を紐付けて登録し、かつ、紐付けされた情報を検索するためのAPIを提供する
- ユーザーが空間参照系<sup>注2</sup>を登録・更新・削除する機能を有している  
(サービスに必要なPI項目を選択できる)

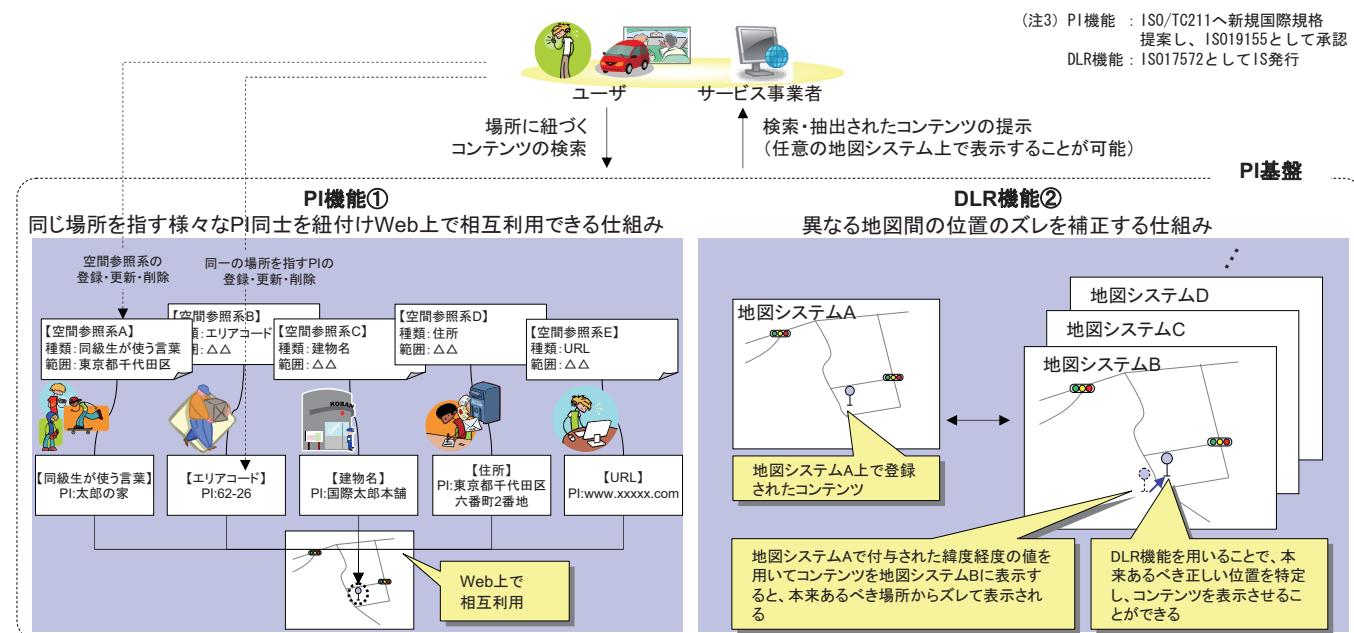
(注2) PI作成上のルール(作成するPIの種類や地理的な範囲などを「空間参照系」として定義することで、自由にPIを作成できる

##### DLR機能

…②

- 異なる地図システム間で発生する位置のズレを補正し、正しい位置を特定する
- 異なる地図システム間で、道路の特徴(形状・名称・交差状況等)をマッチングさせることにより、マッチングした道路からの相対的な位置関係を計算し、地図システム上の対応する場所を高い精度(ヒット率95%以上)で特定することができる
- 携帯端末(スマートフォン等)での利用も可能である\*

##### PI機能、DLR機能ともにISO(国際標準化機構)で定められた仕様に基づいて開発されている<sup>(注3)</sup>



\*印は2009年度の改良・追加技術

igvp-technology@meti.go.jp

## 4.2.1 情報大航海プロジェクト

ユーザインターフェース  
技術

次世代Webサービス

### マッピングサーチ(視覚化検索)

探しているものとは別の価値あるものが見つかる(セレンディピティ)

#### 技術の概要

- 多数のユーザの興味や嗜好(ユーザ行動履歴)を反映した属性をアイテムにベクトルとして付与し、アイテム間の関連の強さを高次元空間中の距離として定義。アイテム間の距離を計算し、それをばね理論により2次元(平面)上にマッピングする
- 行動履歴を元に関連性を算出し、関係性を保った形で2次元にマッピングすることで、ユーザの興味が嗜好性(方向)と距離(近さ)を保ったかたちで表現され、直感的にアイテム間の関係が理解できる
- アイテムの属性を使用せず、ユーザの嗜好性のみで自動的にマッピングを行うため汎用性が高い

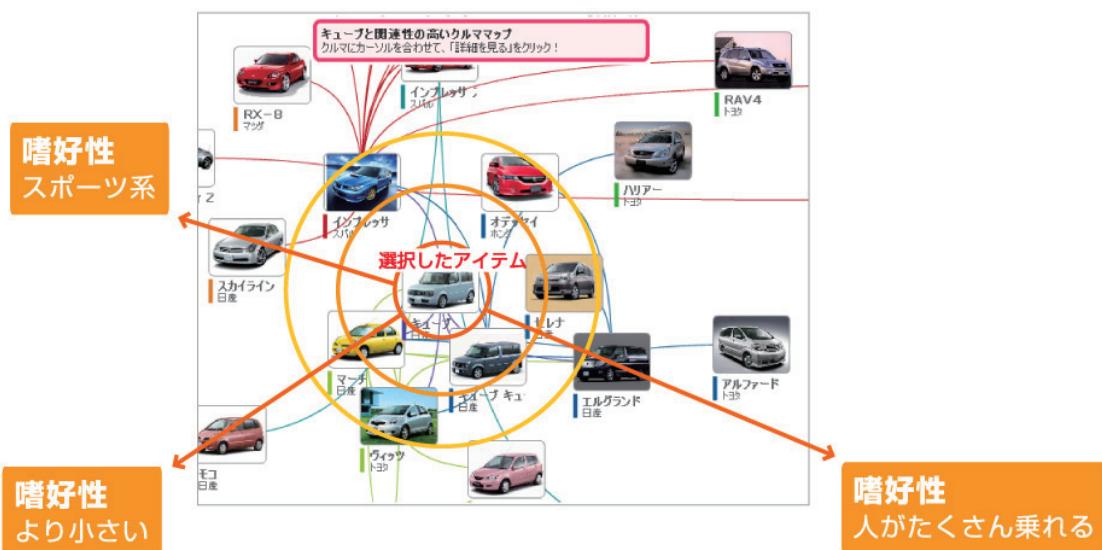
#### 技術の特長

- マッピングを検索に応用したこと、直感的にアイテム間の関係性が理解でき、結果的に探しているものとは別の価値あるものが見つかる
- 2次元マップなので把握しやすい(モニタは2次元であるので2次元に描画した方が操作しやすい)
- キーワードに頼らない検索を可能にするため、キーワード入力に不向きなデバイス、携帯などへの応用も可能
- アイテムの属性を必要とせず、ユーザ行動履歴から自動で描画することで、毎日最新の情報で自動更新されるマップを描画できる
- 関係性を高次元空間中の距離として表現できるものであれば、何にでも適用可能

##### 【例:車を対象としたマッピング】

下図で選択したアイテムを中心にマップされた関係性を人間の目で読み解くと、選択したアイテムに対して右側に行くほど大型の車が表示される傾向があり、スポーツ系の車が間逆の方向性に表示される

このような図を、アイテムの属性なしで描くことができる



igvp-technology@meti.go.jp

## 4.2.1 情報大航海プロジェクト

データマイニング技術

ソーシャルサービス

### リスクモデリング & シミュレーション

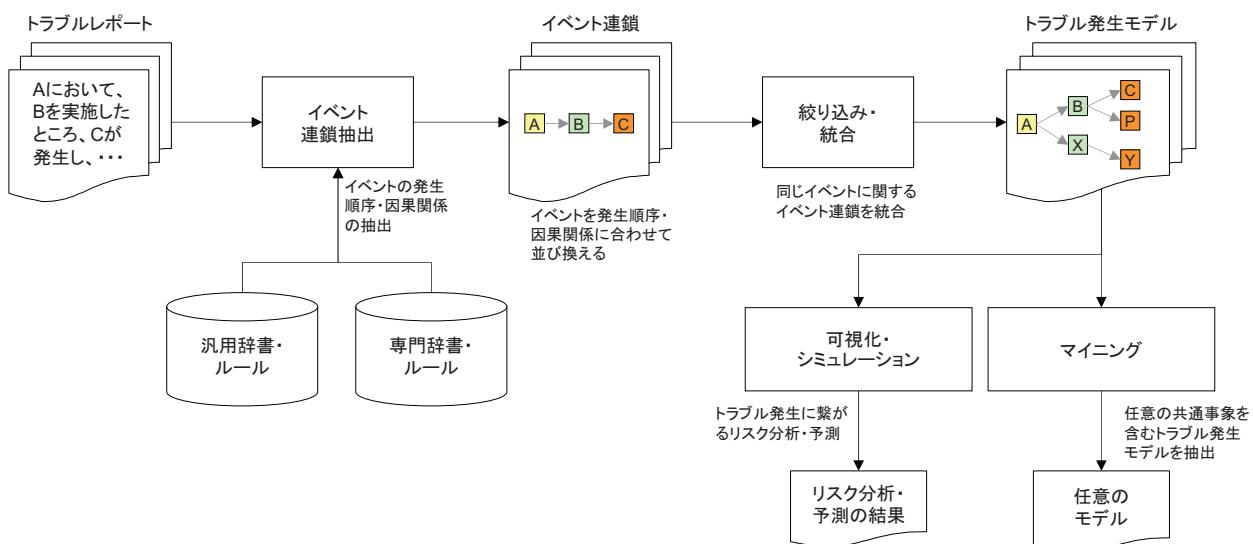
膨大なトラブルレポートからトラブル発生パターンを自動的にモデル化  
リスク予測と予防的対策立案(Proactive Safety Management)を支援

#### 技術の概要

- 自然文で書かれた膨大なトラブルレポートから、イベント(事象、出来事)の連鎖を自動抽出し、得られたイベント連鎖群を統合することにより、トラブル発生モデルを自動構築する
- トラブル発生モデルのグラフ構造に着目して、頻出イベント連鎖パターンを網羅的に探索・抽出することにより、有用な知識の発見を支援する
- トラブル発生に至るまでの経過をモデル化することにより、定量的なリスク分析・予測・シミュレーションを実現し、トラブルに対する効率的な対策を講じることを可能とする

#### 技術の特長

- 経過情報に着目したトラブル発生モデルの自動構築
  - 人手による作業と比較して大量なデータを効率的に扱うことができるため、統計的なトラブル発生モデルを容易に作成することが可能
  - 従来行われていたリスク分析・予測を効率化すると共に、「経験則」の追認や定量化、新たな気づきの発見を促し、より効果的な予防策立案に貢献
- 辞書・ルールベースのイベント連鎖抽出
  - イベント抽出のための辞書やルールを分野依存部分と分野非依存部分に切り分けることにより、他分野への適用性を向上(多分野対応)
  - イベント抽出のための辞書やルールを言語依存部分と言語非依存部分に切り分けることにより、他言語への適用性を向上(グローバル対応)
- トラブル発生モデルのグラフ構造に着目した半自動的知識抽出
  - 複数の事例に共通して現れる頻出イベント連鎖パターンを網羅的に探索・抽出することが可能



## 4.2.1 情報大航海プロジェクト

レコメンデーション技術

ソーシャルサービス

### 治療計画立案エンジン

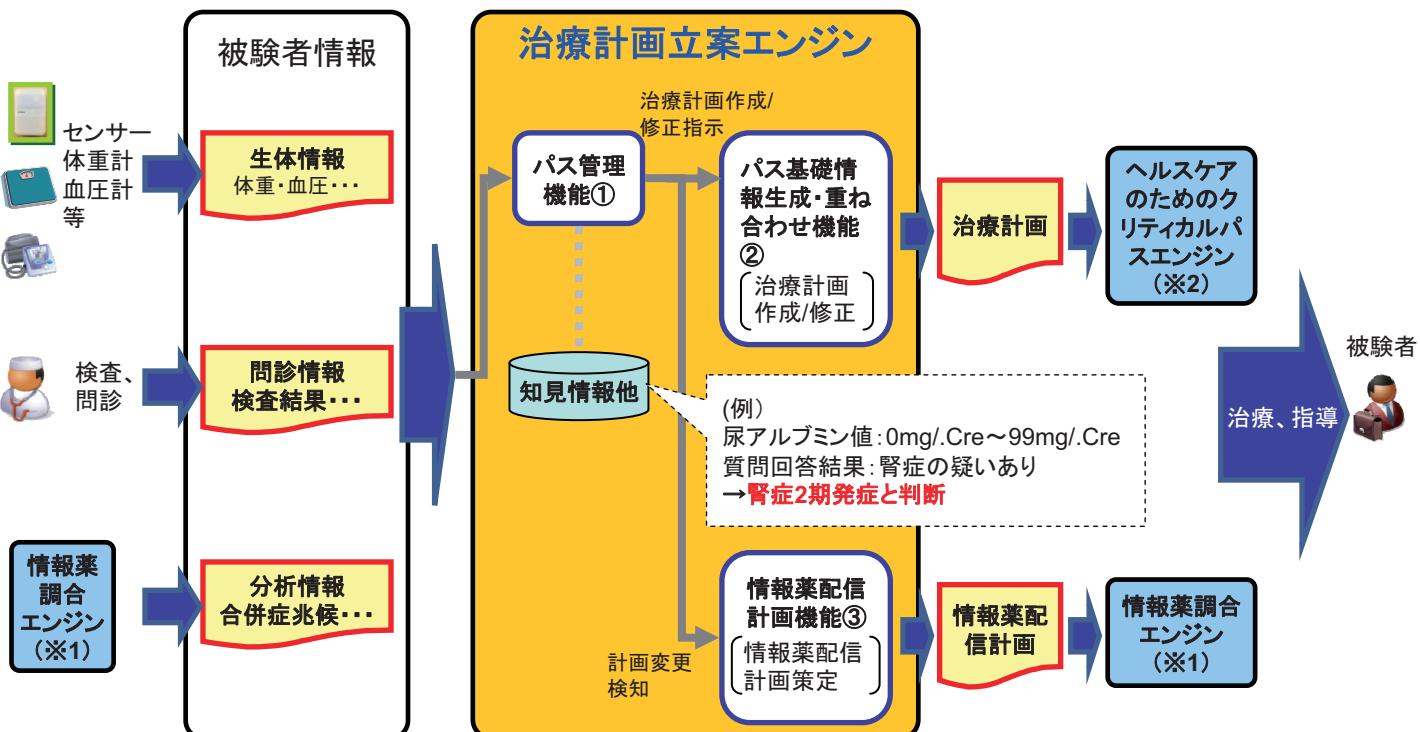
被験者の状態変化を正確に把握し、ガイドラインに準拠した治療計画を自動的に生成

#### 技術の概要

- 被験者ごとの生体情報・検査結果や合併症の兆候を分析した情報を基に、病症変化の状態を判断し、診療ガイドラインや専門医師の知見によるアルゴリズム等を用いて、新たな治療計画や情報薬の配信計画を自動的に策定する
- 専門医の知見やガイドラインに準拠した治療計画案を自動的に提示することにより、非専門医であっても、一定レベルの治療を新たな負担なしに行うことができ、医師の負担軽減や治療品質の向上が実現可能となる
- 本技術は、糖尿病に限らず、様々な慢性疾患医療分野でも適用可能であり、現在手作業で実施している治療計画立案の効率化やサービスレベルの標準化のみならず、新規概念である情報薬配信計画立案が図れる

#### 技術の特長

- パス管理機能 …①  
➢ 生体情報や問診検査結果等、被験者情報と照合し、病症変化の状態を判断するとともに、必要に応じて、パス基礎情報生成・重ね合わせ機能に治療計画作成/修正の指示を行う(パス=治療計画)
- パス基礎情報生成・重ね合わせ機能 …②  
➢ パス管理機能からの指示に基づき、治療計画の作成/修正を行う
- 情報薬配信計画機能 …③  
➢ 治療計画の変更に伴い、新たな情報薬配信計画を策定する



## 4.2.1 情報大航海プロジェクト

レコメンデーション技術

ソーシャルサービス

### 情報薬調合エンジン

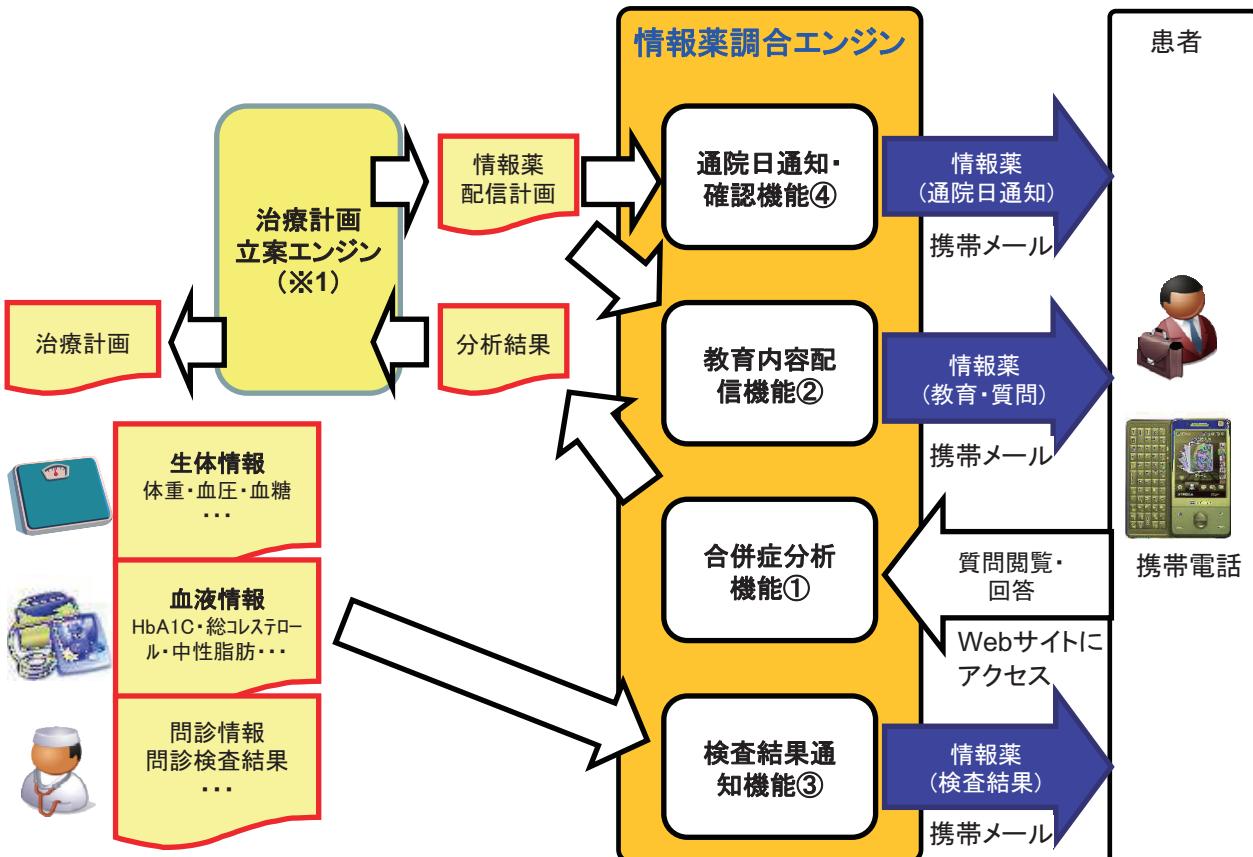
医療機関と患者からの情報を基に患者の個別性に合わせて情報薬を作成

#### 技術の概要

- 医療機関からの検査結果・診断結果と、患者からの生活習慣(行動結果)、バイタルサイン、自覚症状を基に、患者の個別性に合わせて情報薬を作成する
- 情報薬を製造する過程が、医療現場で爆発している情報を収集・検索・整理することを可能とし、医師の負担軽減につながる
- また、患者と医療現場における、情報薬をきっかけとした様々な情報の適時的確な受発信により、病気進行の兆候を早期に発見できるとともに、患者の病気に対する姿勢の改善が見込まれる

#### 技術の特長

- |                |                       |    |
|----------------|-----------------------|----|
| ● 「合併症分析機能」    | : 患者からの回答を基に、患者の状態を判断 | …① |
| ● 「教育内容配信機能」   | : 治療計画に基づき、教育内容を自動配信  | …② |
| ● 「検査結果通知機能」   | : 検査結果を即座に通知          | …③ |
| ● 「通院日通知・確認機能」 | : 次回通院日を自動的に通知        | …④ |



## 4.2.1 情報大航海プロジェクト

自然言語解析技術

次世代Webサービス

### 意味の似ている言葉の抽出

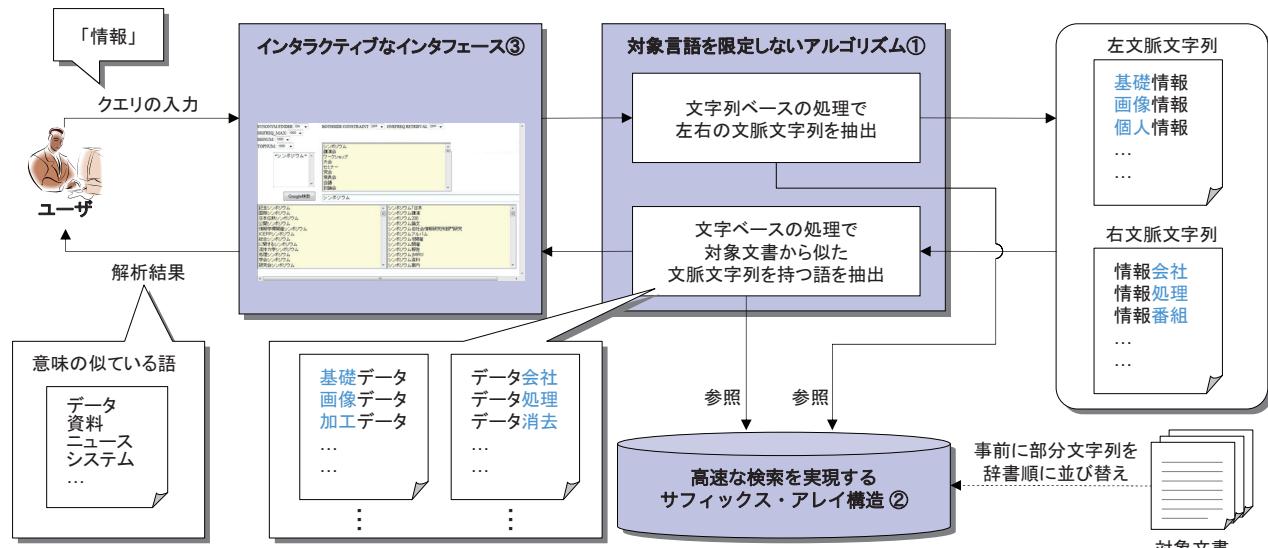
あらゆる分野の文書から意味の似ている言葉(類似語・同義語)の抽出を支援し、レポートなど、大量の文書の分析の効率化・精度向上に貢献

#### 技術の概要

- ユーザが検索する語句(クエリ)を入力後、レポート中からクエリの前後に出現する文字列を抽出し、抽出された文字列を適切な長さで区切った上で、同じ文字列をまとめ、クエリの「文脈文字列」としてユーザへ提示する
- クエリの「文脈文字列」と似た「文脈文字列」を持つ語をレポート中から高速に検索し、類似度順にユーザにクエリと「意味の似ている語」として提示する
- ユーザは提示された「文脈文字列」や「意味の似ている語」を必要に応じて新たなクエリとして選択・検索するなどして、インタラクティブ(双方向)に意味の似ている言葉の抽出を行うことで、様々な分野において、大量にある文書の分析の効率化・精度向上を実現することが可能になる

#### 技術の特長

- 対象言語を限定しないアルゴリズム\*
  - 言語依存の文章解析技術を用いず、文字単位で処理するため、動作原理上、対象言語を限定しない(コーパス中の文字列の出現パターンに基づいて、文字列を適切な長さに区切る)...①
- 高速な検索を実現する「サフィックス・アレイ構造」を利用した文書事前処理\*
  - 対象文書内の部分文字列を辞書順に並び替えておくことで、高速な検索を実現する
  - 一般的なクエリに対しては、既存手法に匹敵する同義語抽出の精度を担保しつつ、数秒で同義語を抽出する...②
- インタラクティブなインターフェース\*
  - 抽出した意味の似ている語を保存し、必要に応じて新たなクエリとして入力するための効果的なインターフェースを実装している
  - 「サフィックス・アレイ構造」によって応答速度の速い快適な操作性を実現している...③



\*印は2008年度の改良・追加技術

## 4.2.1 情報大航海プロジェクト

時空間情報解析技術

ソーシャルサービス

### センサ情報に基づく行動解析基盤

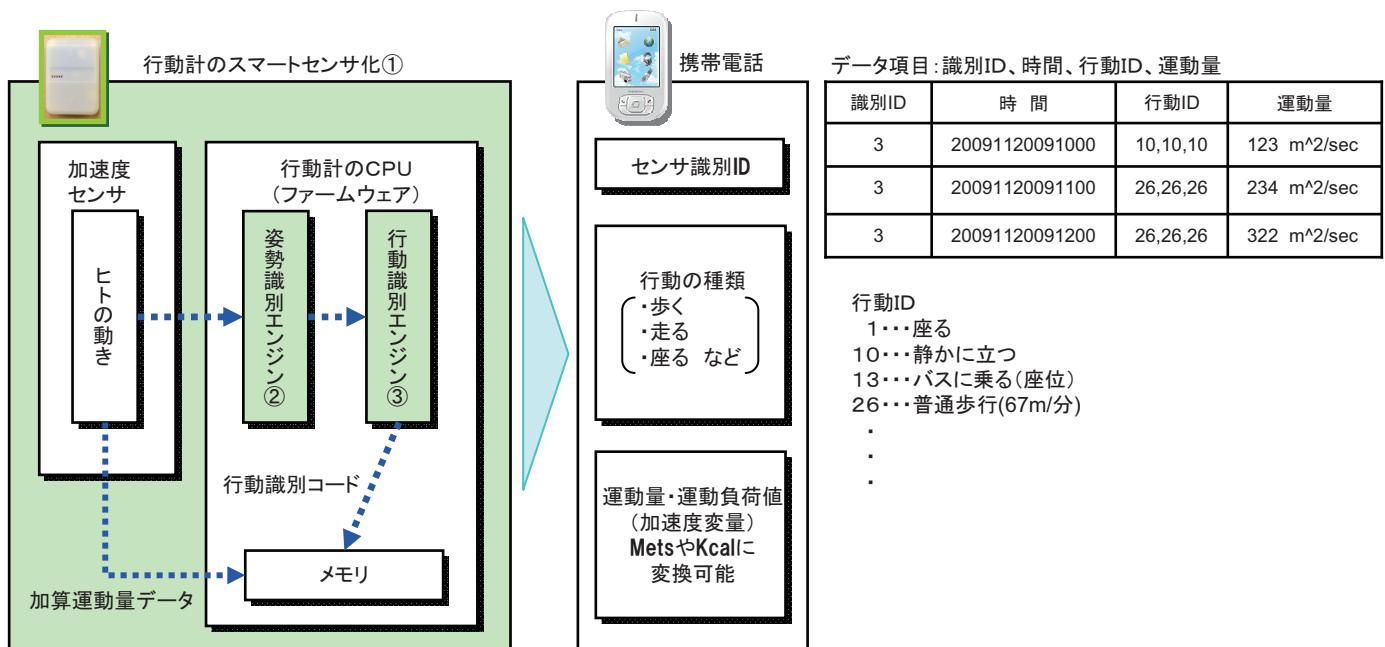
対象者の行動情報を自動的に収集、解析できるプラットフォーム

#### 技術の概要

- 3軸加速度センサ情報を用いて、行動計内で利用者の行動を自動識別する
- 今後加速度センサ情報の他に、GPS、ローカル位置センサ、体温計等の様々なデバイスからの情報を加えることが可能となる基盤として開発しており、位置情報等も含め更に精緻に利用者の行動を識別することが可能となる
- 本技術を、健康指導に活用することにより、利用者の行動等に即した健康指導を自動生成し、携帯情報端末に配信するサービスを実現できる

#### 技術の特長

- 「行動計のスマートセンサ化」
  - 行動計に高性能CPUを搭載し、行動計内で行動識別を行うとともに運動量も計算する \*…①
- 「姿勢識別エンジン」
  - 年齢、体重、性別などの利用者情報と、一定期間毎の重力加速度（利用者の腰に装着した3軸加速度センサから得られる上下、前後、左右方向の加速度データ）を解析することにより、姿勢について立位、横臥（左右）、うつ伏せ等の姿勢を識別する…②
- 「行動識別エンジン」
  - 3軸各方向の加速度データの時間的变化から、行動について静止、乗車等の行動を識別する…③



\*印は2009年度の改良・追加技術

## 4.2.1 情報大航海プロジェクト

ユーザインターフェース  
技術

ソーシャルサービス

### 健康管理自己情報コントロール技術

#### 健康管理を行うための汎用的な情報収集・配信システム基盤

##### 技術の概要

- 個人が自分自身の健康管理を実施するためには、日々の定期的な計測結果、医療機関における検査結果を保存、チェックし、個人で能動的に管理することが理想的である
- これまで、日々の定期的な計測結果(例えば、体重計、血圧計)や医療機関における検査結果(血液検査結果)は、記帳を行ったり、検査結果を待つ必要があった。加えて、疾病を患っている場合には、疾病によっては患者にもその疾病に関する専門知識が必要となり、基礎的な知識や注意すべき事項の周知が課題となっていた
- 本技術はこうした課題を解決するため、日々の計測結果や医療機関における検査結果を取得保存するデータベース、各個人の疾病に対して適したコンテンツ(検査項目、検査項目のトレンド)を配信するコンテンツ配信エンジンを提供し、個人で健康管理を実施する環境を提供する。本技術は、さまざまな疾病、利用形態を想定した設計を行い、健康管理を行うための汎用的な情報収集・配信システム基盤としての利活用が期待できる

##### 技術の特長

###### ● 種々疾病への対応が可能

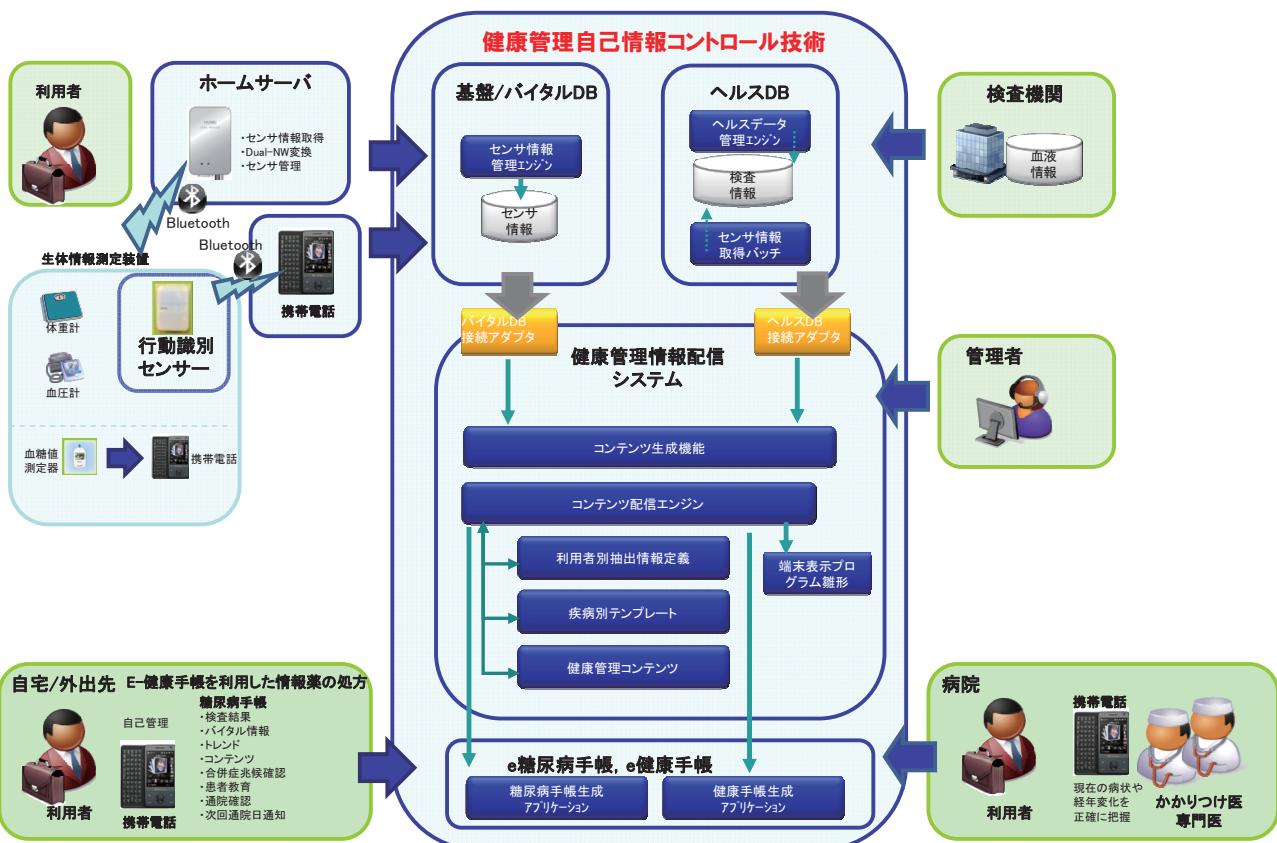
疾病別テンプレートを疾病毎に準備することで、種々疾病に対応することが可能

###### ● 利用者個人がカスタマイズ可能

利用者個人が健康管理のために、参照する情報をカスタマイズし、トレンド情報や判定値を確認することが可能

###### ● 検査機関や計測機器からの情報の取得が可能

医療機関における検査結果や、計測機器(体重計等)の計測結果を取得し、保存することが可能



## 4.2.1 情報大航海プロジェクト

データマイニング技術

パーソナルサービス

### 行動情報マイニングエンジン

移動ログ情報から個人の行動特性や場の特性を見出す汎用の解析機能を提供

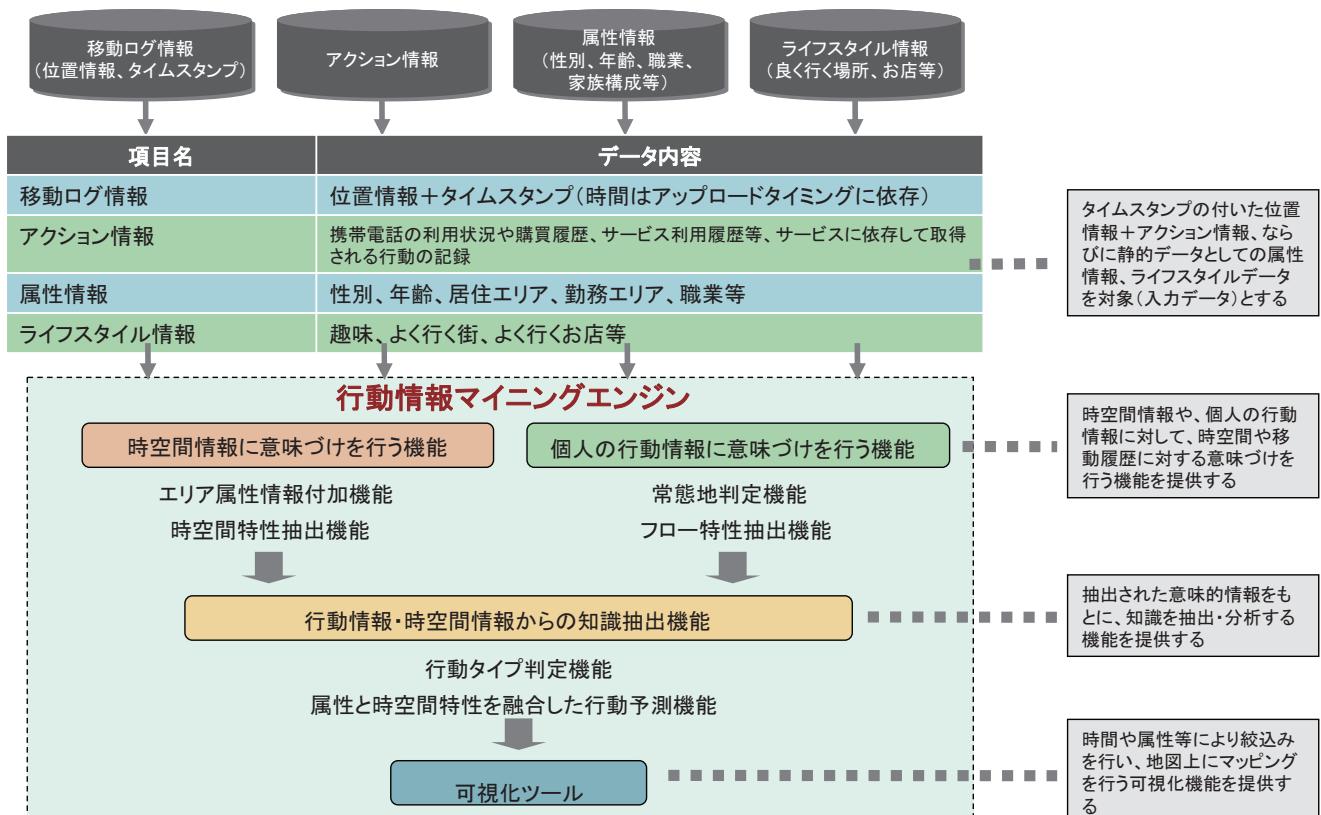
#### 技術の概要

- 携帯電話のGPS機能等により収集される個人の移動ログ情報(位置情報+タイムスタンプ)から、個人の行動特性や、場の特性を解析するための共通基盤技術(ツール群)を提供する
- 具体的には、移動ログ情報を解析することにより、時空間に対する意味付けを行う機能、個人の行動履歴に対して意味付けを行う機能、ならびにこれら意味付けされた情報から意味のある知識の抽出機能を提供する
- 本技術は、顧客の行動や状況を把握するデバイスとして今後利活用が進むと思われるGPS端末やその他の位置情報収集デバイスを用いた行動ログ(※)データに対する汎用的な解析基盤を提供する

※行動ログ=移動ログ情報+アクション情報(携帯端末の利用履歴や購買情報等)

#### 全体構成

- 本技術は、入力データとして、移動ログ情報(位置情報+タイムスタンプ)、アクション情報、属性情報、ライフスタイル情報から構成されるデータを前提とする
- 本技術は、大きく以下の4機能から構成される
  - 時空間情報に意味付けを行う機能(場所と時間帯に対して、それがどのような特徴を持った場所であるか)
  - 個人の行動情報に意味付けを行う機能(個人の行動ログから、各地点の持つ意味を解析)
  - 行動情報・時空間情報からの知識抽出機能(意味付けされた移動ログから意味のある知識を抽出する機能)
  - 可視化ツール(解析結果を条件を指定することにより地図上に表示させる機能)



## 4.2.1 情報大航海プロジェクト

データマイニング技術

パーソナルサービス

### グループ属性・エリア情報抽出技術

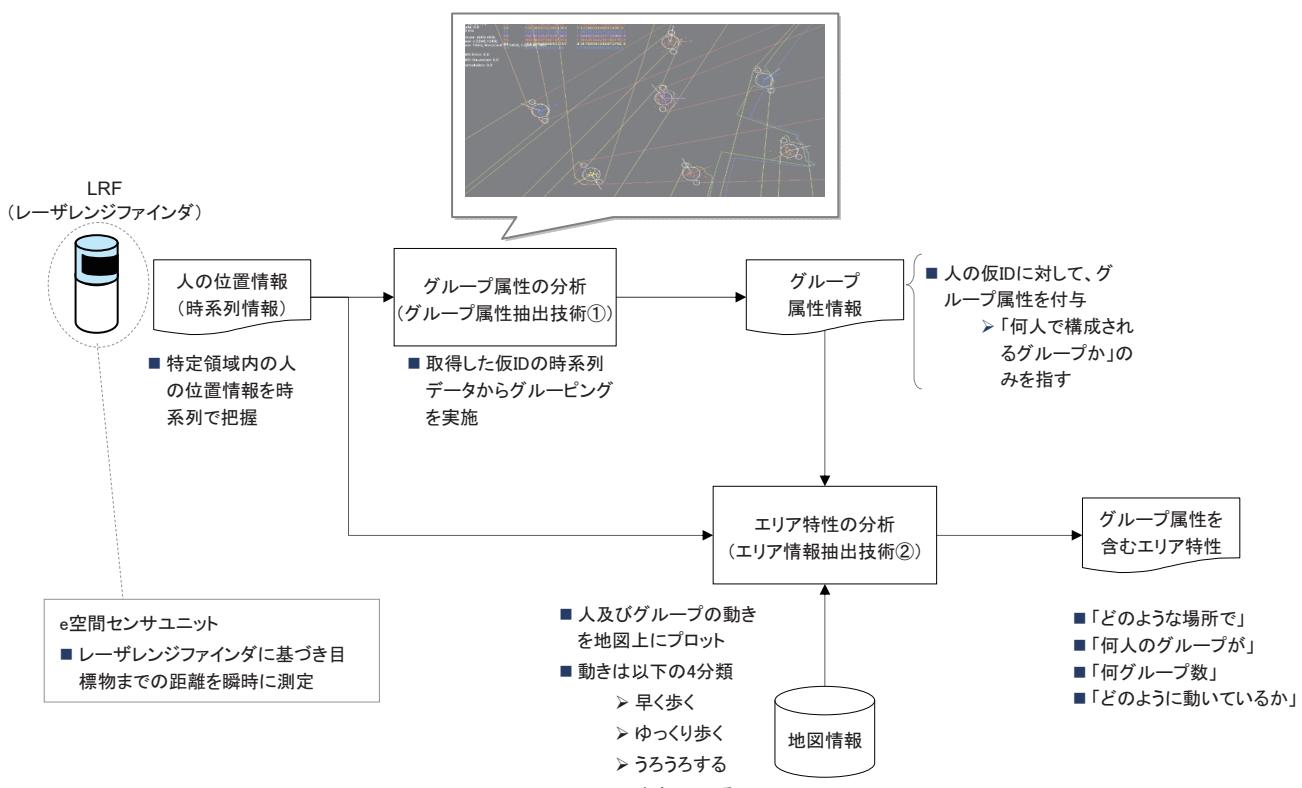
空間内を移動する人々のグループと、空間の利用のされ方を、「見える化」できます

#### 技術の概要

- 個人情報の利用を極力抑えた形で、特定の空間内における不特定の人々のグループ属性、及びグループの空間内エリアでの動き方を分析する
- 従来、複合商業施設や店舗等において、施設の使われ方を把握するにはアンケートや口コミ情報等の情報源に依存せざるを得なかったため、店舗側は画一的なレコメンド情報を提供することに留まっていた
- それに対し、個人や団体などのグループによって施設が実際どのように利用されているかが把握可能となることによって、来訪者のプロファイルを事前に入手することなく、グループ属性に応じたレコメンドの提供が可能となる

#### 技術の特長

- グループ属性抽出技術 ①
  - 特定空間における人の位置情報を時系列で把握することが可能である
  - 各IDの位置情報の変化(時系列情報)からグルーピングすることが可能である
  - 従来は連続した位置追跡が困難であった人同士が近距離ですれ違う場合や人の行動が複雑になる店舗内においてもグルーピングが可能である\*
- エリア情報抽出技術 ②
  - 特定空間内をエリアに区分した上で、属性が付与されたグループの移動状況から、各エリアの特性を把握することが可能である



\*印は2009年度の改良・追加技術

## 4.2.1 情報大航海プロジェクト

レコメンデーション技術

パーソナルサービス

### 複数の推薦エンジンの選択・組合せが可能なマルチモード推薦

ユーザーの行動コンテキストに基づき複数の子推薦エンジンを組み合わせる  
推薦技術

#### 技術の概要

- ユーザーの行動コンテキストを基に利用者がどのようなコンテンツが欲しいかを示す利用者モードを推定し、利用者モードを基に子推薦を選択・集約してコンテンツを配信する、コンテキスト適応マルチモード推薦基盤
- 推定した利用者モードをユーザーに公開し、さらにユーザーによる利用者モードの変更を許し、その変更内容に基づいて推薦することを可能とする利用者モード推定・可視化・制御機能

#### 技術の特長

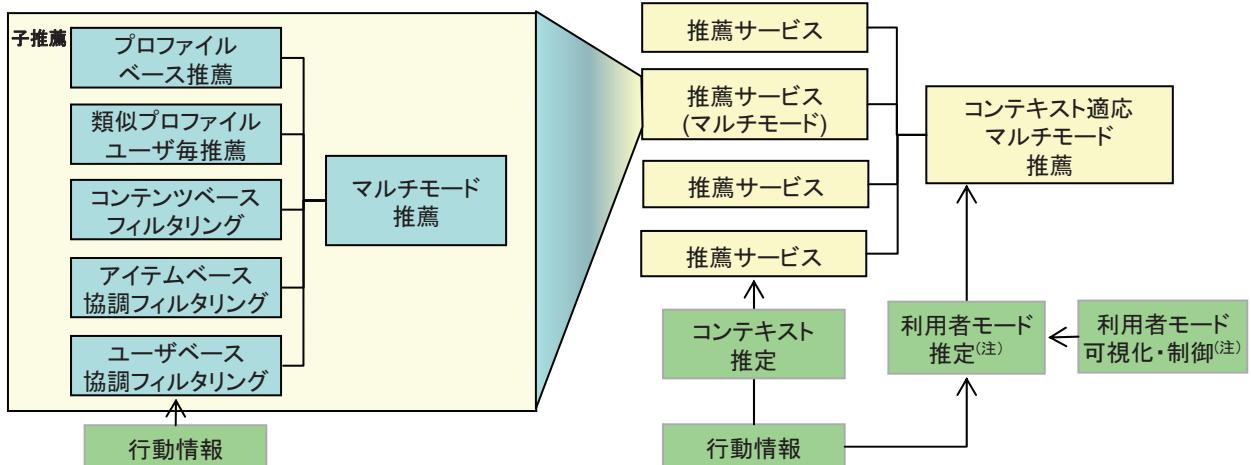
- コンテキスト適応マルチモード推薦基盤
  - ユーザーの位置(現在地や次に滞在すると予測される位置)や位置属性(自宅や会社等)、ユーザーのプロファイル等のコンテキストに応じて子推薦を選択し、子推薦の推薦候補を集約・配信できるプラットフォーム
  - 外部開発の推薦サービスを子推薦として組み込める推薦エンジンオープン連携機能
- 利用者モード推定・可視化・制御機能
  - コンテキスト適応マルチモード推薦基盤で取得した子推薦ごとの推薦結果から、コンテキスト毎に各子推薦の重み分布を学習し、学習結果に基づいた推薦を可能とする利用者モード推定機能
  - 学習した利用者モードの重み分布を公開し、ユーザによる重みの制御を受け入れた推薦を可能とする利用者モード可視化・制御機能

平成20年度

平成21年度

複数の推薦方式を個人に合わせて選択・集約・学習するマルチモード推薦。一部子推薦が行動情報を利用

行動情報からコンテキスト推定と利用モード推定を行い、推薦サービスを選択・集約・学習するマルチモード推薦



(注)期待される応用分野(その1)参照

## 4.2.1 情報大航海プロジェクト

レコメンデーション技術

パーソナルサービス

### 実世界の行動とネット上の行動を統合したユーザ特性推定

多様な行動ログに基づくコンテキスト推定技術と受容されやすい推薦情報の推定技術

#### 技術の概要

- 実世界での行動ログと、ネット上での行動ログを収集し、それらの情報に基づいてユーザの行動の特性を推定する
- 実証サービスにおいて開発・実現される行動ログの収集機能によって情報を蓄積し、それを解析することによってユーザの行動時のコンテキスト(ユーザの置かれている状況)と、各ユーザに適した推薦情報を推定する
- これまで別々に存在していた実世界の行動情報(お買い物など)と、ネット上の行動情報(ブログの投稿など)を統合・整理してユーザの興味や意図を推定することで、ユーザにとって価値の高い情報を提供するサービスを実現することが可能となる

#### 技術の特長

##### ● 心的コンテキスト推定\*

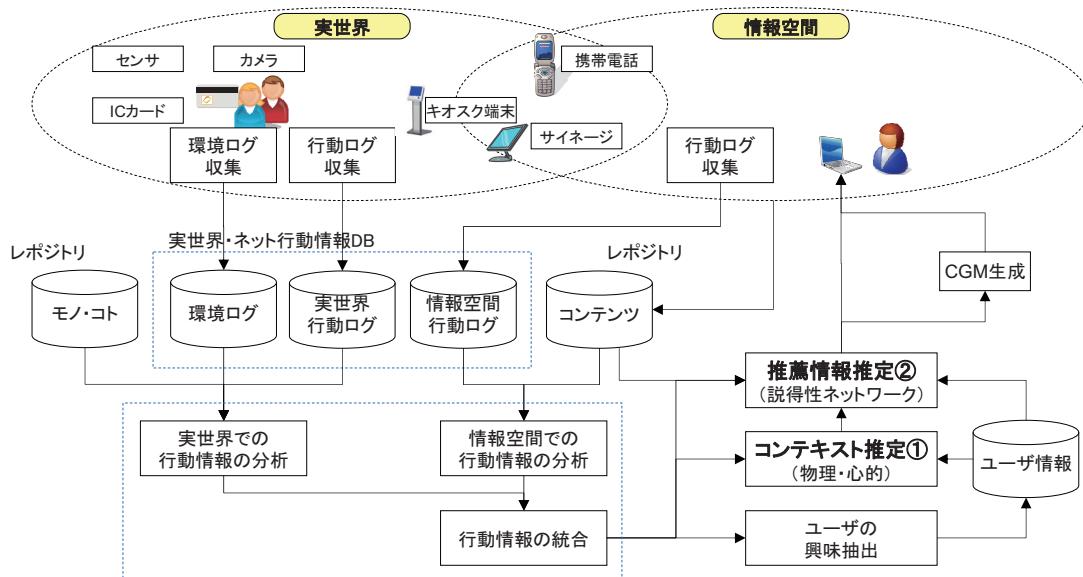
…①

- 従来技術では、コンテキストの推定は、主にユーザがいる場所や時間などの外的要因から行っていたが、本技術では、ユーザの目的や感情などの内的要因まで考慮することができる
- 実世界の行動情報とネット上の行動情報を、どちらか片方だけではなく、両方の情報を統合してコンテキストを推定することが可能である

##### ● ユーザが受け入れやすい推薦情報を配信可能にする「説得性推定」\*

…②

- 従来の協調型の情報推薦では、「似たユーザは同じ情報を望む(同じ情報が役に立つ)」という前提の下、ユーザ同士の類似性に基づいて推薦情報が決定されてたが、本技術では、「説得が行われる際に情報の送り手が有する特性(説得性)によって、同じ情報であっても、その受容度(情報の受け入れられやすさ)が変化する」という社会心理学に基づいて類似性以外の要因も考慮することを可能にしている
- 説得性のモデルを魅力度(好意、ユーザ間の類似性)と信憑性(専門性、信用性)から構成し、これらの要素を実世界とネット上で収集した行動情報を用いて推定・算出することで、ユーザが受容しうる推薦情報推定を実現することができる



\*印は2008年度の改良・追加技術

igvp-technology@meti.go.jp

## 4.2.1 情報大航海プロジェクト

プライバシー情報管理  
技術

パーソナルサービス

### プライバシ情報セキュア流通基盤

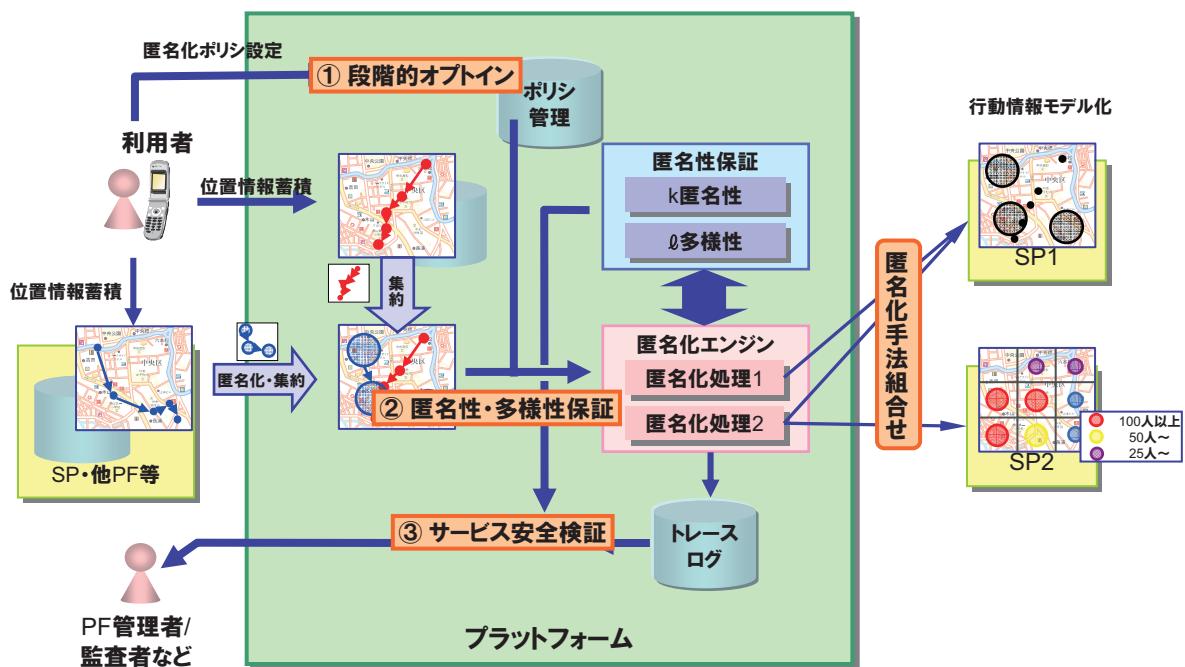
#### 安全・安心な利用者行動情報利活用のためのプラットフォームの実現

##### 技術の概要

- 利用者の行動情報の匿名性を確保しつつ、安全に利活用する流通基盤として、プライバシー情報の匿名化により安全に利用者の行動情報を二次利用できる情報配信を確立し、安心・安全なプラットフォーム（マイ・ライフ・サービスプラットフォーム）を実現する
- プラットフォーム（PF）では、利用者が安心して行動情報を提供できるよう、利用者主導の匿名性の制御として、利用者の行動情報を匿名化する際の方針となる匿名化ポリシーを利用者が指定する手段、さらに、利用者が自身の行動情報の蓄積状況やサービスプロバイダ（SP）への提供状況を確認しながら、匿名化ポリシーの変更や提供後の削除を行う手段を提供する
- SPへ提供された行動情報の匿名性が利用者が指定する匿名化ポリシーを満たしているか、実際に評価し検証する

##### 技術の特長

- 匿名度可視化による段階的オプトイン …①
  - 利用者毎に設定可能な匿名化ポリシーを導入し、各設定における提供情報を可視化することで、段階的に提供情報を調整し、享受するサービスと提供情報のバランスを取りやすくすることを実現
- PF・SP連携における匿名性・多様性保証 …②
  - PF・SPから提供可能なプライバシー情報を集約し、行動情報モデル化の要件に応じて匿名化手法を組み合わせ、k匿名性やl多様性等の指標を満たす匿名化処理を実現
- 匿名性評価によるサービス安全性検証 …③
  - プライバシー情報を処理するプログラムから呼び出されるログ記録ライブラリが蓄積するトレースログに基づいて、プライバシー情報漏えい検査と提供情報の匿名性評価を行う



## 4.2.1 情報大航海プロジェクト

プライバシー情報管理技術

パーソナルサービス

### パーソナル情報保護・解析基盤

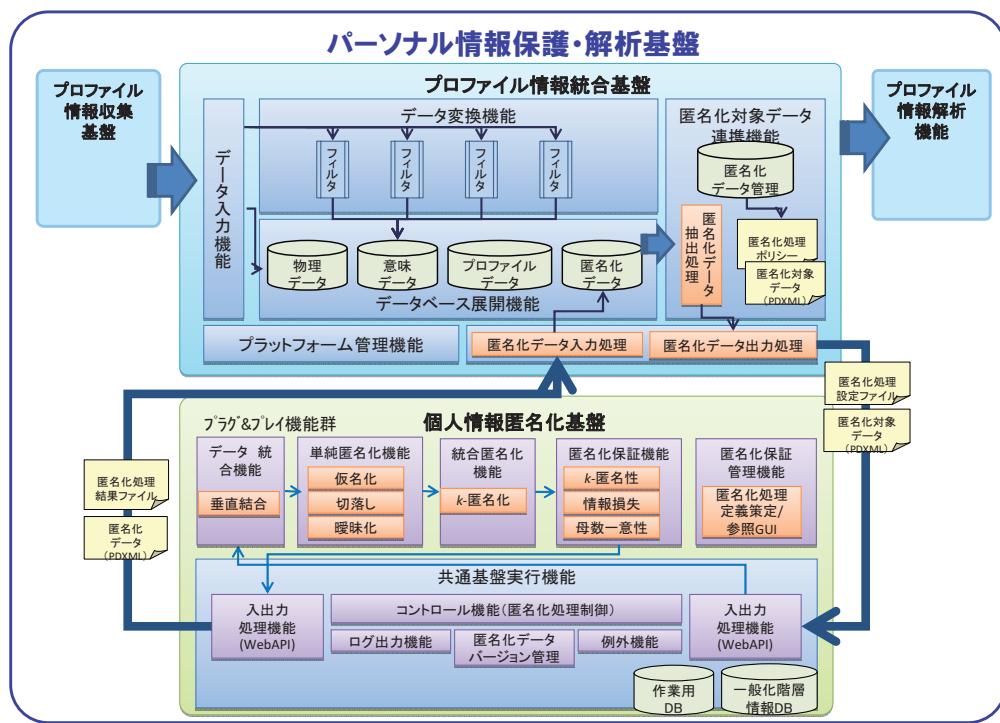
パーソナル情報を利活用し社会的価値を持たせるための基盤共通技術

#### 技術の概要

- パーソナル情報(生存する特定の個人を識別する「個人情報」だけでなく、個人の行動や購買履歴等個人と連結可能な情報の総称)を収集・統合・解析・匿名化するためのプラットフォームを提供する
  - 本技術により、利活用目的に応じた適切な匿名化処理が行われたことを保証(安全性を担保)するので、パーソナル情報を安心して利用できる
  - 本技術の活用によって、個々人の顕在意識・潜在意識の喚起につながるパーソナライゼーションサービスなどを実現し、新たな産業の発展を促す枠組みとなることを目指す
- ※ 現在の個人情報に関する法制度は、匿名化技術による情報利用に対応しきれていないため、本技術の開発と並行して制度課題の検討も進行中である

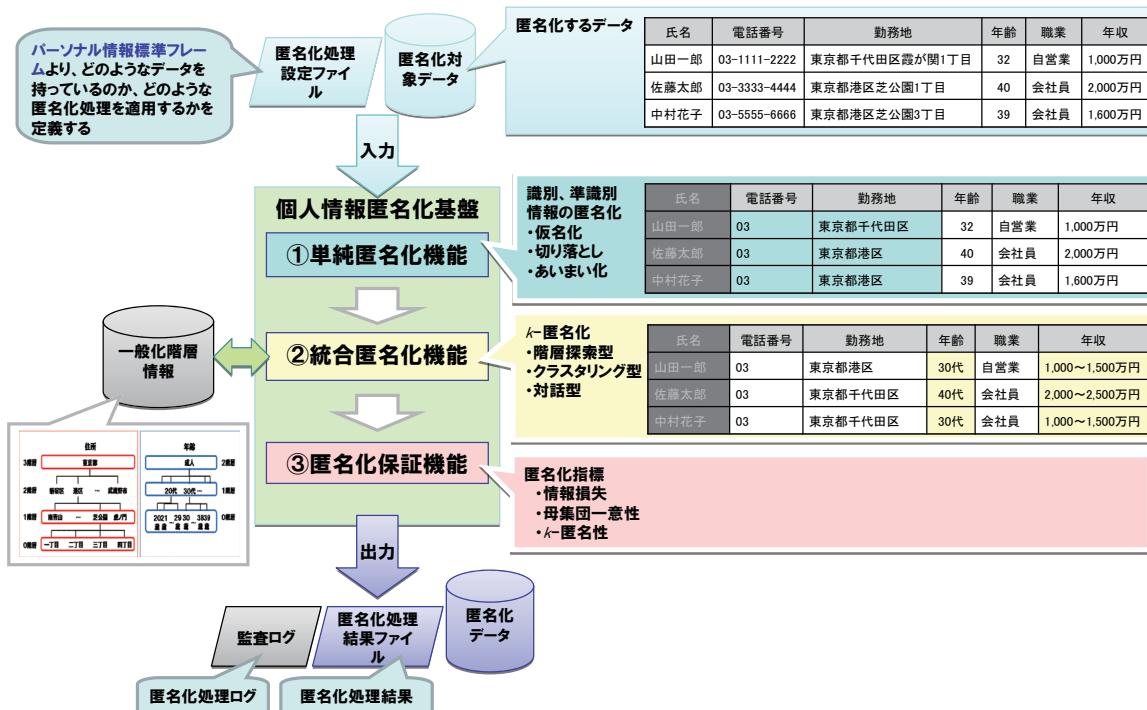
#### 全体構成

- 本技術は、大きく「個人情報匿名化基盤」、「プロファイル情報統合基盤」の2つの基盤から構成され、両基盤技術を連携して利用することによりパーソナル情報の統合的な利活用プラットフォームを実現する。両基盤は、それぞれを単独で利用することも可能である
  - 個人情報匿名化基盤
    - ◆ 「単純匿名化機能」、「統合匿名化機能」、「匿名化保証機能」、「匿名化保証管理機能」、「データ統合機能」、「共通基盤実行機能」から構成される、パーソナル情報を匿名化し二次利用を可能とするプラットフォーム
  - プロファイル情報統合基盤
    - ◆ 「プロファイル情報収集基盤」、「プロファイル情報解析機能」と組合せることで、個人の行動を把握するための各種センサ情報を収集・統合・解析するプラットフォーム



## 技術の特長(その1)【個人情報匿名化基盤の機能と処理の流れ】

- 個人情報匿名化基盤は、収集したパーソナル情報に対し、個人が特定されないように匿名化を行い、匿名化したデータの安全性を評価できる匿名性の指標を算出する
- そのため、匿名化対象データに対し、①単純匿名化(識別情報の仮名化・切り落とし、準識別情報のあいまい化)、②統合匿名化( $k$ -匿名化)、③匿名化保証(匿名性、有用性の評価)の処理を適用し、匿名化データを生成する



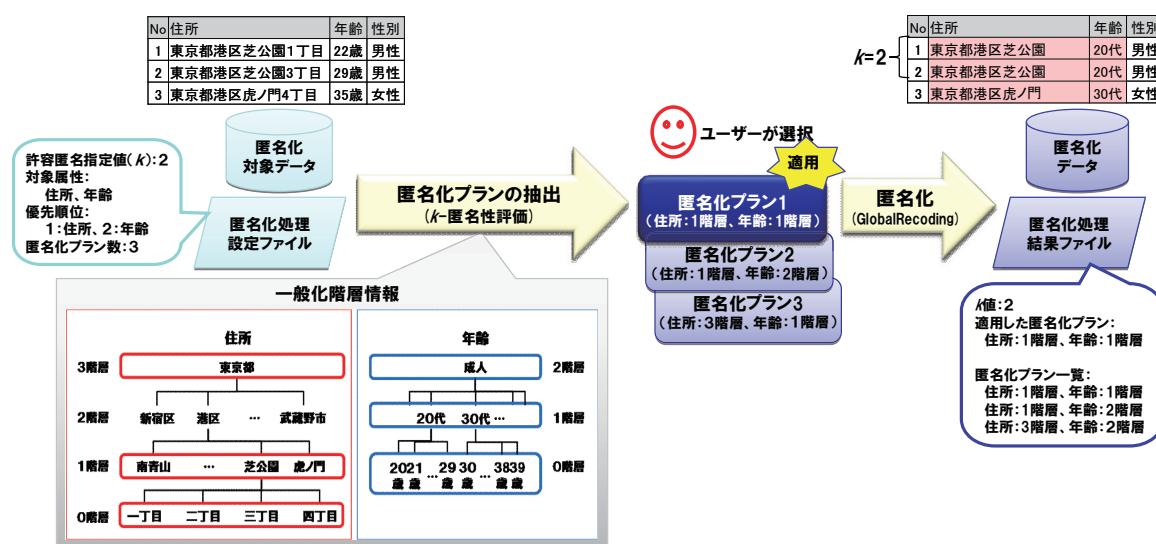
## 技術の特長(その2)【統合匿名化機能】

### $k$ -匿名化を実現している

- $k$ -匿名化とは、属性情報の一般化や抑制などにより、 $k$ -匿名性を充たすように共通の属性情報の組み合わせを持つ複数のレコード集合を構成することである
- 階層探索型、クラスタリング型、対話型の3方式から、利用場面に応じた選択が可能である

### 階層探索型の例

優先順位が高い準識別情報( $k$ -匿名化の対象となる属性)から順に、各属性の一般化階層情報を用いて匿名化した後の $k$ -匿名性を評価し、許容匿名指定値( $k$ )で指定された $k$ -匿名性を満たす複数の匿名化プラン( $k$ -匿名化の候補)を抽出し、匿名化を施す



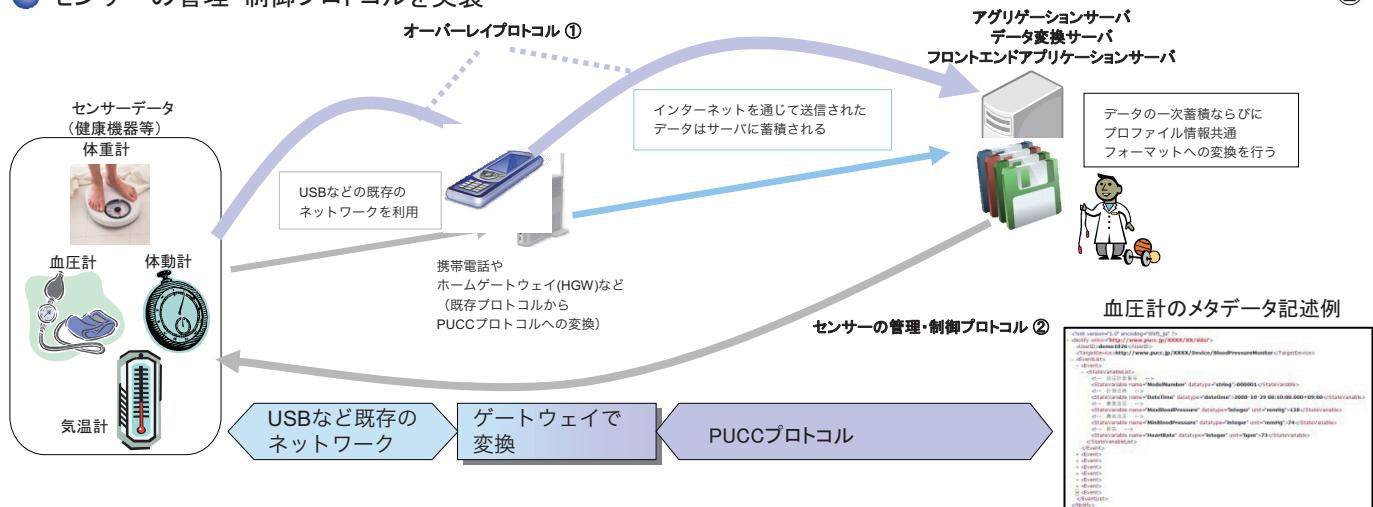
## 技術の特長(その3)【プロファイル情報収集基盤】

- 異なる通信プロトコル間での相互接続を可能にする「オーバーレイプロトコル」
    - オープンオーバーレイ通信とメタデータ標準を中心としたデバイス通信プロトコル(PUCC<sup>(注)</sup>プロトコル)を採用
    - AVネットワーク、センサーネットワーク、インターネット、モバイルネットワーク、赤外線通信などを相互に接続することが可能
    - 各種センサーデータ形式を包含するPUCCメタデータ仕様を採用し、新しい通信方式や新デジタル機器の出現に対しても柔軟かつ簡単に接続可能
    - グローバルオープンスタンダード技術として、様々な国際標準化団体と共同作業を推進

1

- センサーの管理・制御プロトコルを実装

2



(注) PUCC: P2P(Peer to Peer) Universal Computing Consortium

## 技術の特長(その4)【プロファイル情報解析技術】

- 「IGVminer (LCM<sup>(注1)</sup>、VSOP<sup>(注2)</sup>等を含む)」

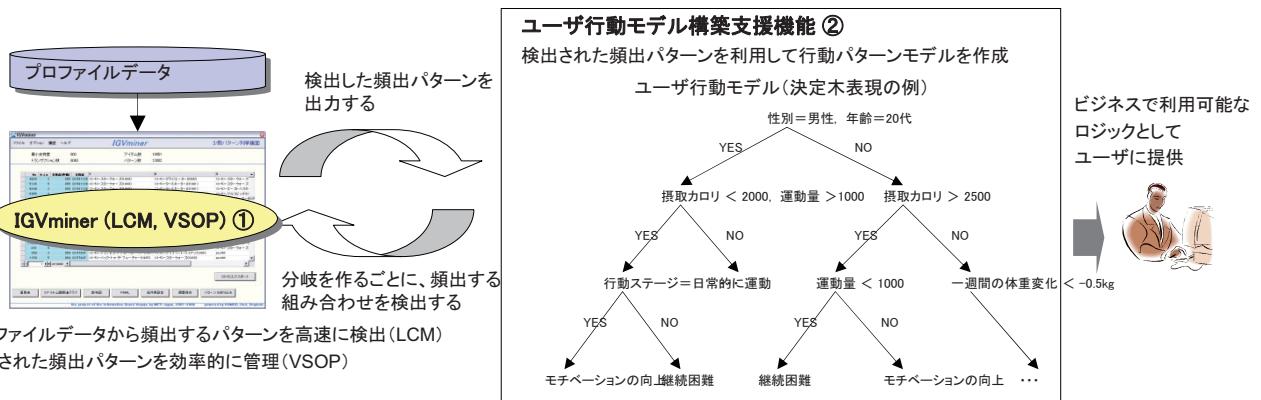
1

- 1,000万件を超える大量の行動履歴データや時系列データから、一定の頻度以上で出現するパターン（頻出パターン）を秒速1万個を超える世界最高水準のスピードで発見することが可能
  - データに重みをつけることで、特定のユーザーや状況にのみ頻出するパターンを優先的に検出
  - 時系列を含んだデータからイベントの順序関係も考慮した頻出パターンの検出が可能であることに加え、時系列を適切に考慮することを可能にするデータ変換プログラムを提供（サービス共通技術「頻出パターンマイニング」：（株）ジー・サーチ）
  - 検出された頻出パターンを効率的に管理できる仕組みを提供

- ## ● 「ユーザ行動モデル構築支援機能」

• • • 2

- ▶ 検出された頻出パターンから、個人の嗜好や行動、あるいは環境特性等を予測するモデルを作成し、決定木表現等で可視化。またタクソノミ機能や、シーケンシャルパターンと属性データの組み合わせからの頻出パターン生成機能も有する



(注1) LCM: Linear time Closed itemset Miner (国立情報学研究所 宇野毅明准教授) (注2) VSOP: Valued-Sum-of-Products(北海道大学 湊真一准教授)

## 4.2.2 情報信憑性検証技術

情報の信頼性・信憑性  
検証技術

情報の信憑性・信頼性情  
報を提供するWebサービス

### Webコンテンツの分析技術（画像・音声・映像）

Web上のマルチメディア情報の信憑性判定の支援を行います

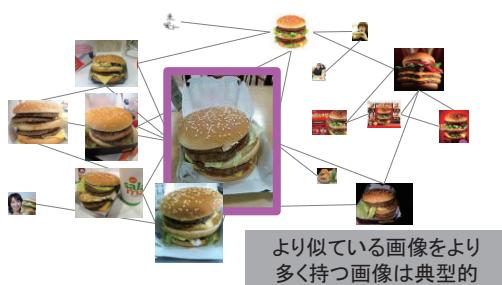
#### 技術の概要

- 画像、音声、映像などの情報とそれを説明するテキスト情報や周辺のコンテンツから、コンテンツの内容を推定し、Web上から同等のテキスト情報とそれと関連づけられた別のマルチメディア情報を取得する
- 対象とする情報と取得された情報を分類するなどして、マルチメディア情報とそれを説明するテキスト情報の整合性・不整合性を分析する
- Web上の画像、音声、映像を含む情報の信憑性を評価する手がかりを情報利用者に提示することによって、画像、音声、映像情報とそれを説明するテキスト情報の不整合度を検証することができるようになる

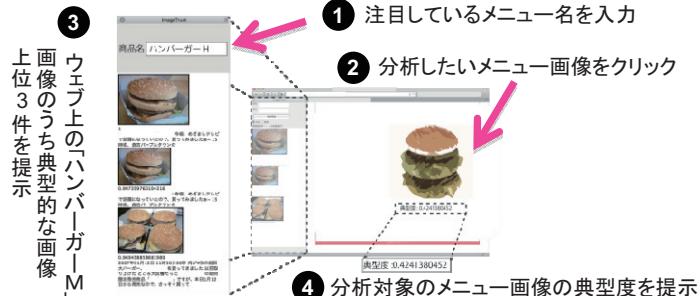
#### 技術の特長

##### 画像間の関係性分析に基づく画像信憑性分析

目標：ある商品名に関する画像の典型度を算出



##### リアルタイム画像信憑性分析システム ImageTrust

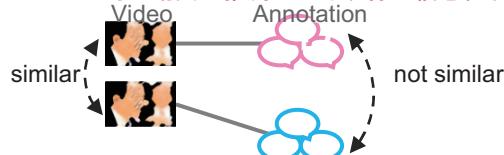


#### 映像の整合性・不整合性分析

ステークホルダに対する記述・表現の不整合を分析



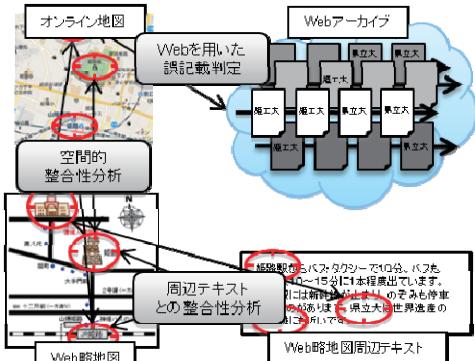
#### アノテーション対立動画の推薦による映像の信憑性判断支援



同じ映像を含み、アノテーションが対立している動画は、元の動画に対して異なる編集が施されている可能性がある

ある映像と映像に対するアノテーションを入力として与えた際  
映像が類似／アノテーションが非類似な動画を提示する

#### 実空間への要求意図に基づく地理的整合性分析



## 4.2.2 情報信憑性検証技術

情報の信頼性・信憑性  
検証技術

情報の信憑性・信頼性情  
報を提供するWebサービス

### Webコンテンツの分析技術（テキスト情報）

Web上のテキスト情報の信憑性判定の支援を行います

#### 技術の概要

- Webコンテンツを構成するテキスト表現の表層的特徴(文体、表現、語彙等)やテキスト情報の発信者の専門性等、Webコンテンツの信憑性を評価する手がかりとなる側面におけるテキスト情報の分析を行う
- 分析結果を提示することによって、情報利用者がWebコンテンツを構成するテキスト情報の信憑性を評価する基準を与えることができる
- ブログ情報の信憑性・信頼性を判断する方法として、情報発信者(ブロガー)が過去にどのような記事を発信したのかを把握することにより、発信者の信頼性を推定する

#### 技術の特長

サーチエンジンの索引データを用いた  
テキストの表層的特徴分析



##### コンテンツの公正さを表す指標

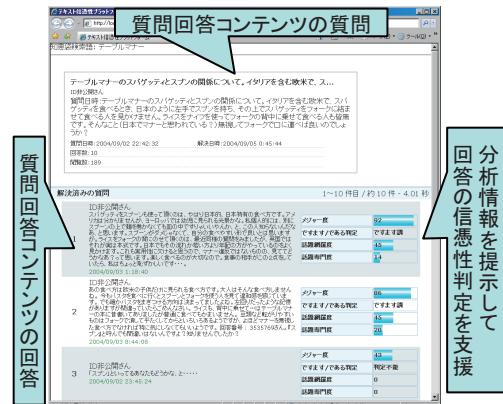
コンテンツそのものの特徴

- ・ メジャー度
- ・ 話題網羅度
- ・ 話題専門度
- ・ 体裁(ですます／である判定)

##### コンテンツの社会的支持度を表す指標

コンテンツの外部からの評価

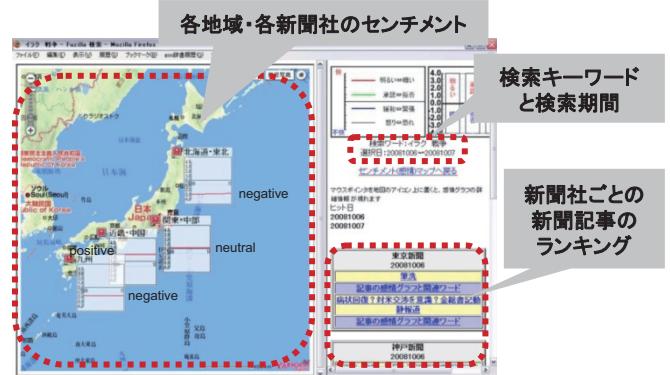
- ・ ソーシャルブックマーク件数
- ・ 披リンク数
- ・ 地域的支持度(広域支持度／近接支持度)



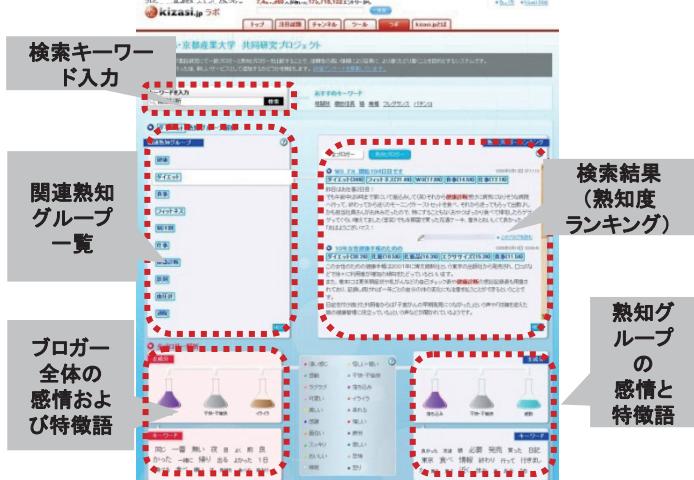
センチメントおよび熟知度分析に基づく  
発信者分析技術

地域性に基づく新聞社の観点差異を可視化することが可能  
熟知度に基づく高信頼性ブログ情報を取得することが可能

#### 1. 新聞記事を対象としたセンチメントマップシステム



#### 2. ブログを対象としたブログランキングシステム



<http://www.dl.kuis.kyoto-u.ac.jp/i-believe/>

#### 4.2.2 情報信憑性検証技術

## 情報の信頼性・信憑性 検証技術

情報の信憑性・信頼性情報を提供するWebサービス

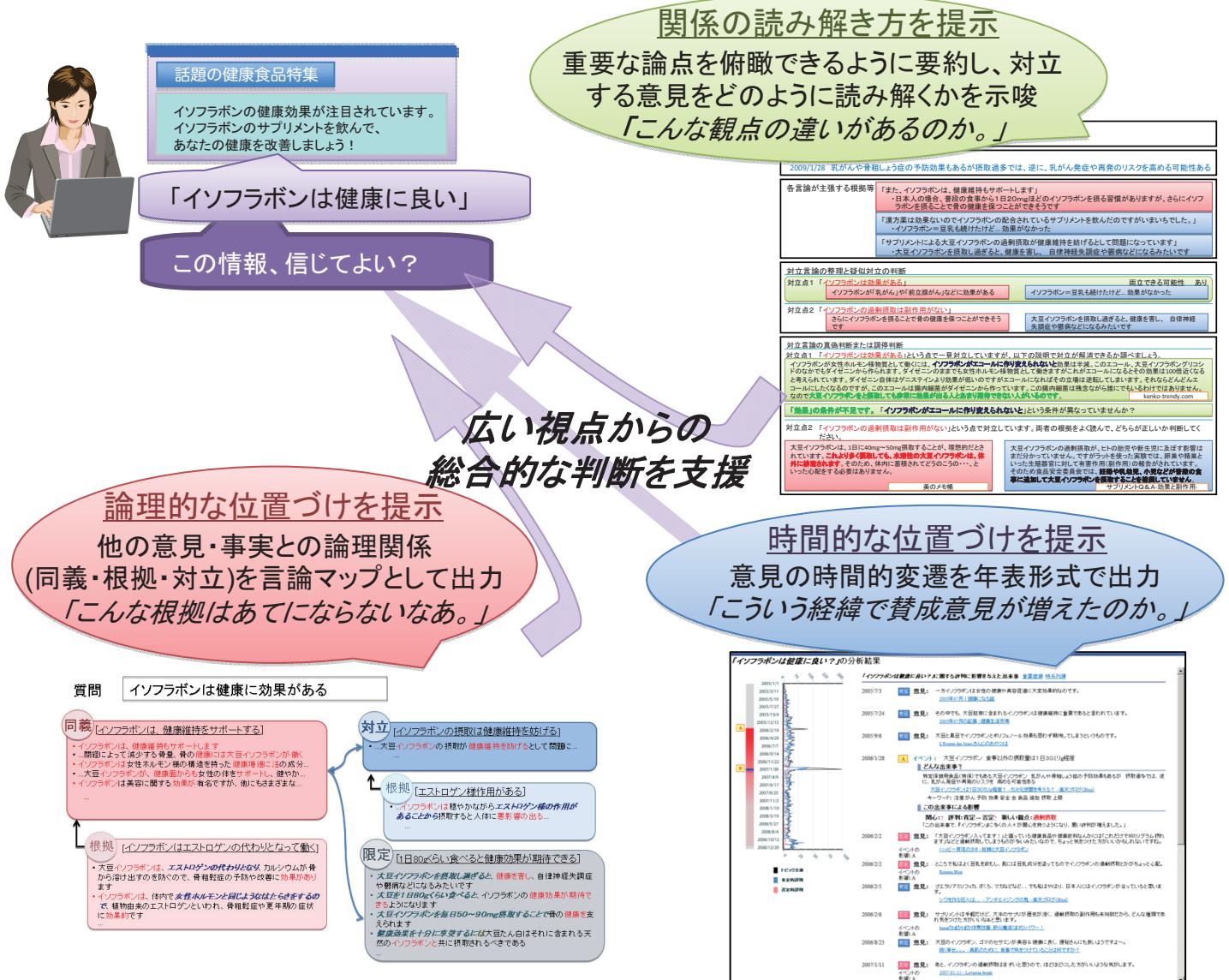
## 意味内容の時系列分析技術

大量のWeb文書から、着目する情報の信憑性判断に有用な情報を抽出・提示

## 技術の概要

- ネットワーク上に流通する玉石混淆な情報の中から、情報利用者が個々の判断基準で信頼できる情報を容易に判断・利活用できる環境を実現する
  - 従来のテキスト検索技術では幅広い意見を偏りなく閲覧することができず、利用者は一面的な情報だけから判断することを強いられていた
  - それに対し、本プロジェクトでは、着目するテキスト情報に論理的・時間的に関連する情報を幅広く収集し、それらの関係を分かり易く提示する技術を開発することで、利用者の情報信憑性判断を支援する

## 技術の特長



## 4.2.2 情報信憑性検証技術

情報の信頼性・信憑性  
検証技術

情報の信憑性・信頼性情  
報を提供するWebサービス

### 【WISDOM】Web情報の信頼性分析技術

#### 玉石混淆のWebコンテンツから信頼できる情報を発見

##### 技術の概要

- 膨大なWeb情報から信頼できる情報を見つけるために、自然言語処理に基づく情報分析手法で、多種多様なWeb情報を発信者、内容、外観で組織化してユーザに提示する
- 従来のWeb検索は人気のある少数のページを閲覧することを目的としており、発信者毎の主要な意見やその分布、少数の対立意見などのWeb情報の全体像を把握することが困難であった
- それに対し、本技術は、検索結果のWebページ集合から、主要な意見や対立する意見を抽出し、さらに発信者やページ種別で分類して俯瞰表示することで、利用者の情報信頼性検証を支援することが可能である

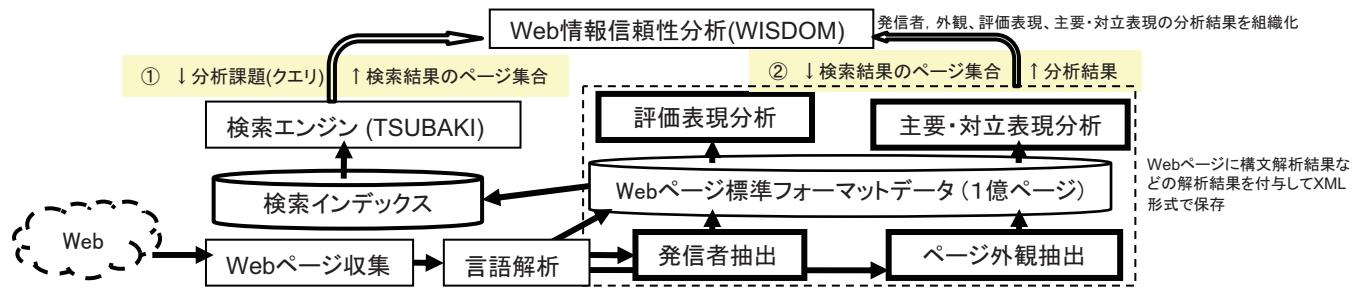
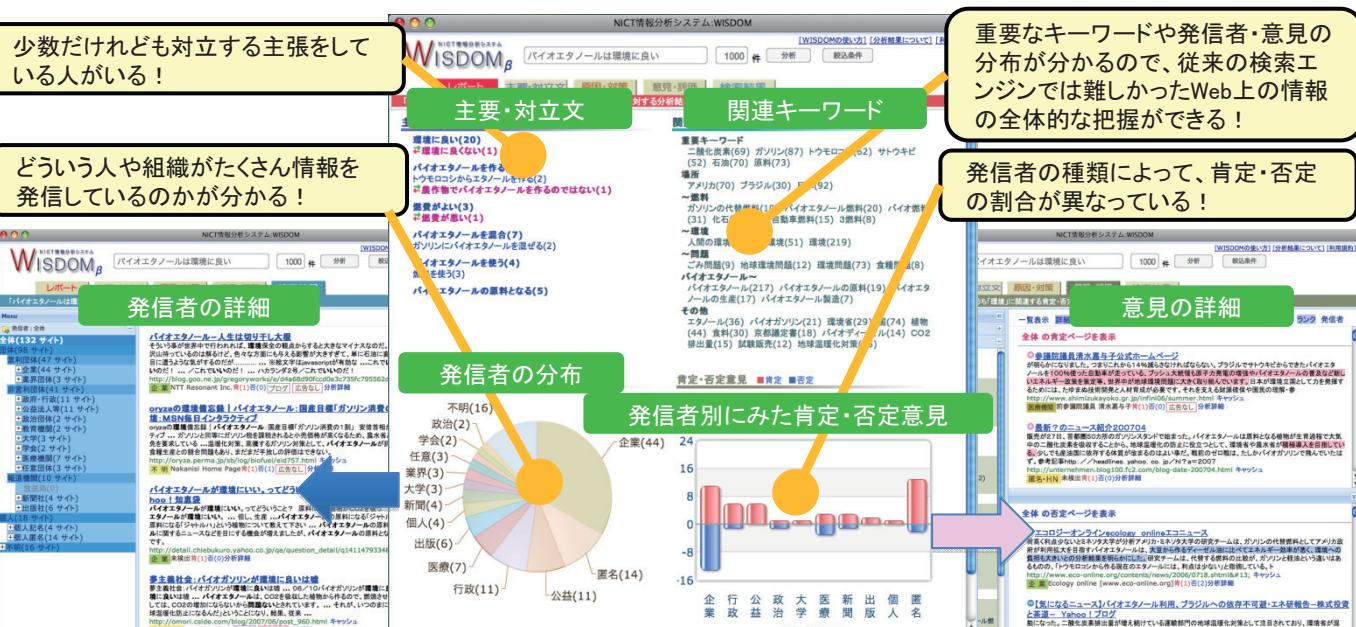
##### 技術の特長

###### テキスト内容分析技術

- 評価表現分析:「商品Aが好き」のような主観的な意見だけでなく、「商品Aは買って3日で壊れた」のような客観的な意見も抽出し、肯定・否定別に分類することが可能
- 主要・対立表現分析:主要な表現だけでなく、少数であっても主要文と対立する表現を抽出することが可能

###### 発信者・外観分析技術

- 発信者分析:Webページから情報発信者を抽出し、発信者クラス毎にWebページや意見を自動分類
- ページ外観分析:Webページをブログなどのページタイプで分類し、さらに連絡先の有無や広告量で分析



<http://kc.nict.go.jp/project1/>

## 4.2.3 超高性能データベース基盤ソフトウェア

データベース基盤技術

DB利用サービス

### 革新的実行原理に基づく超高性能データベース基盤ソフトウェア

「非順序型実行原理」によって解析系データベース処理の飛躍的な高速化を実現

#### 技術の概要

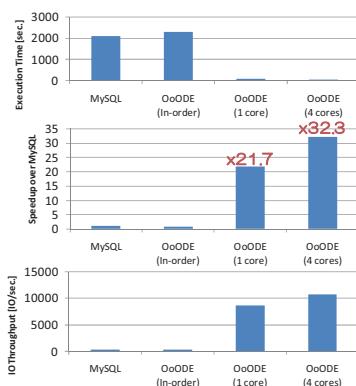
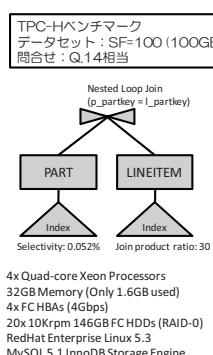
- 情報爆発時代を迎え、ペタバイト級のデータベースが登場しつつある。これら超巨大情報の戦略的利活用によって、新規の高付加価値産業の創出、社会システムの抜本的効率化など、我が国の産業競争力の強化や国民の安心・安全の実現が期待されている
- データベース処理性能の飛躍的な向上が鍵であり、センサーネットワーク情報解析や流通トレーサビリティシステムなど、情報爆発時代における「超巨大情報の戦略的活用」を可能とするために、非順序型実行原理に基づき飛躍的な性能向上を実現する超高性能データベースエンジン（OoODE: Out-of-Order Database Engine）を開発する
- 平成23年度末までに、従来型データベースエンジンと比較して約100倍の解析系処理の高速化を目指す

#### 技術の特長

- 従来型データベースエンジンと非順序型データベースエンジンの実行原理の比較



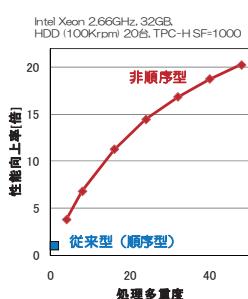
- オープンソースDBMS(MySQL)ベースの実験  
マルチコア環境へに限定的対応による飛躍的高速化



- 商用DBMS(HiRDB)ベースの実験  
複雑度の高い問合せにおける高速化

複雑度の高い問合せにおける性能向上  
(例：指定地域における指定商品種別の市場シェア変化を分析：TPC-H Q.8)

```
SELECT o.year,
       sum(case when nation = 'BRAZIL' then volume else 0 end) / sum(volume) as
       mkt_share
  FROM (
    SELECT year,
           extract(year from o.orderkey) as o.year,
           l.extendedprice * (1 - discount) as volume,
           n2.name as nation
      FROM part,
           supplier,
           lineitem,
           orders,
           customer,
           nation n1,
           nation n2,
           region
     WHERE p.partkey = l.supplierkey
       AND s.supplierkey = l.supplierkey
       AND l.orderkey = o.orderkey
       AND o.custkey = c.custkey
       AND c.nationkey = r.nationkey
       AND r.name = 'AMERICA'
       AND s.nationkey = n2.nationkey
       AND o.orderdate between '1995-01-01'
                           AND '1995-12-31'
       AND p.type = 'ECONOMY ANODIZED STEEL'
       AND p.size < 3
       AND l.quantity > 0
   GROUP BY o.year
  ORDER BY o.year)
```



## 4.2.4 センシングWeb

プライバシ情報管理  
技術

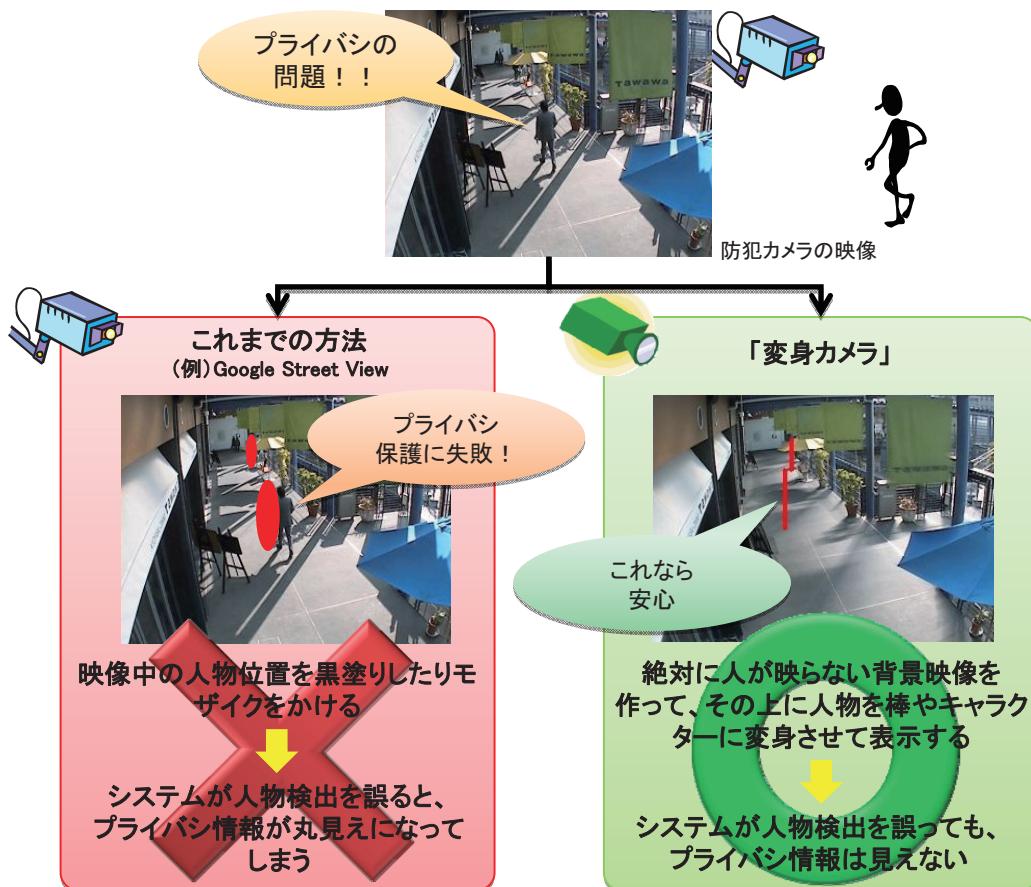
センシングWebサービス

### プライバシ情報の管理

プライバシ侵害のないデータしか出力しない「変身カメラ」・「安心センサ」の実現

#### 技術の概要

- センサの生データに含まれるプライバシ情報を除去する技術
- 信号処理・パターン認識技術の性能が100%正確でなくても、プライバシの侵害が絶対に起こらない処理

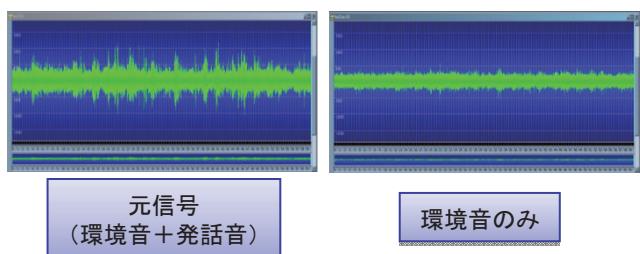


#### 技術の特長

- カメラにおける背景推定と前景抽出処理
  - シーンの不变性・周期性を用いた背景推定
  - シーン適応型の人物検出処理



- マイクにおける環境音推定と発話音抽出処理
  - 多数の発話音付き環境音サンプルの学習に基づく環境音推定
  - 発話音の音質変換・音声認識



## 4.2.4 センシングWeb

プライバシ情報管理  
技術

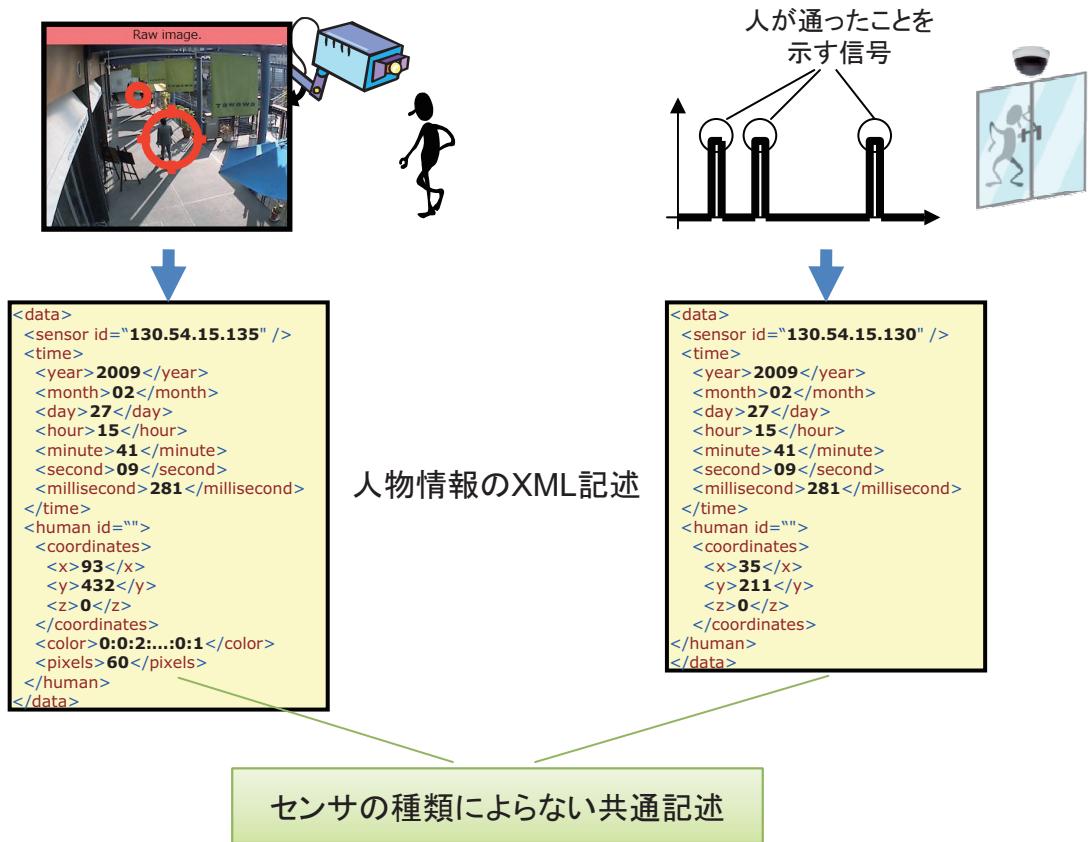
センシングWebサービス

### センサ情報の共有

センサの種類によらない共通XML記述の提案と  
利用者からの情報要求に応じたフォーマット変換技術

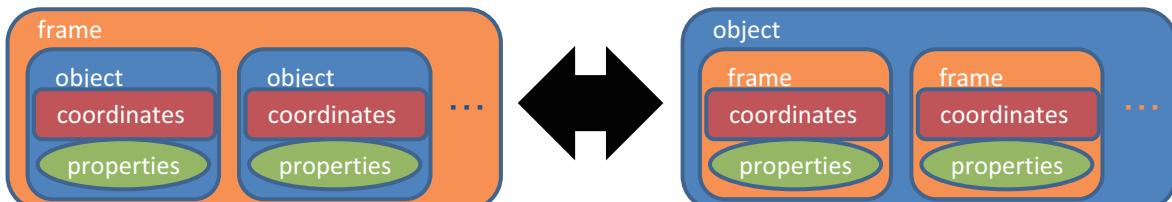
#### 技術の概要

- 利用者やサービスプロバイダがセンサの種類を意識せずに任意のセンサの情報を利活用するための、センサ情報記述の統一化
- センサデータから取得できる情報と利用者側から求められる情報を考慮した記述設計



#### 技術の特長

- 時空間的断面(フレーム)に関するデータ記述と対象ごとのデータ記述との相互変換技術



- 次世代ロボット連携群のインターフェース仕様を基にサービス仕様記述、要求仕様記述

## 4.2.4 センシングWeb

プライバシ情報管理  
技術

センシングWebサービス

### 観測型コンテンツの提示

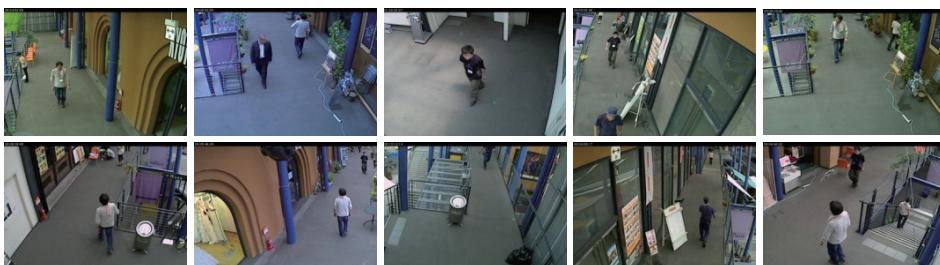
「にぎわいマップ」・「ディジタルジオラマ」・「シースルービジョン」等の開発

#### 技術の概要

- センシングWebにおけるセンサ記述形式に準じたアプリケーションの実装
- 複数のセンサデータを統合した、直感的・効果的な情報提示

#### 技術の特長

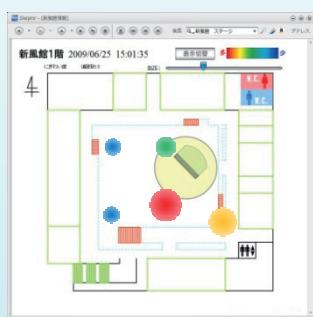
- 実証実験会場「新風館」でのアプリケーション実装



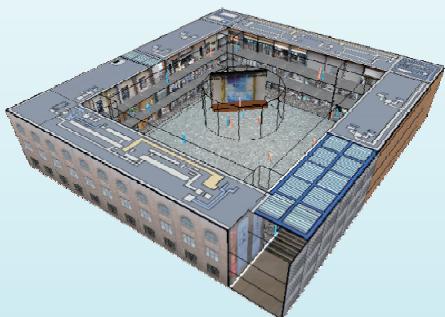
センサの生データをそのまま見ても、その場所の状況が直感的に理解しづらい



#### 俯瞰的な可視化による直感的空間把握



「にぎわいマップ」



「ディジタルジオラマ」

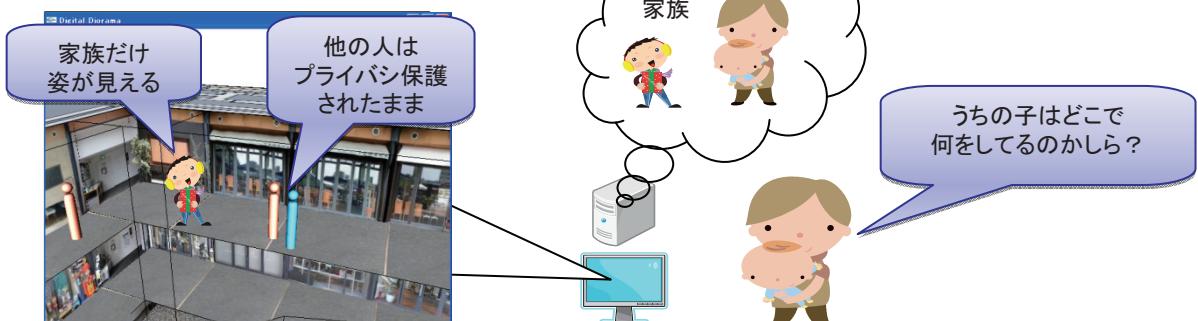
#### ユーザ視点でのセンサデータ可視化



「シースルービジョン」

- 被観測者と閲覧者の関係性を考慮したプライバシ kontrole

- RFIDタグ等を用いた個人同定技術との連携

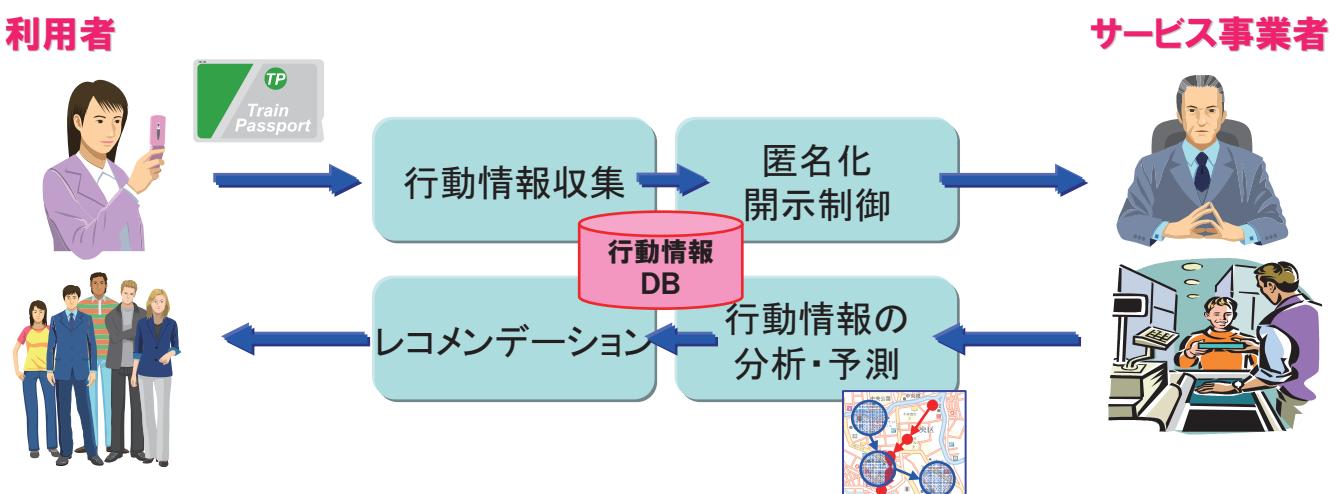


### 4.3 基盤技術の活用例

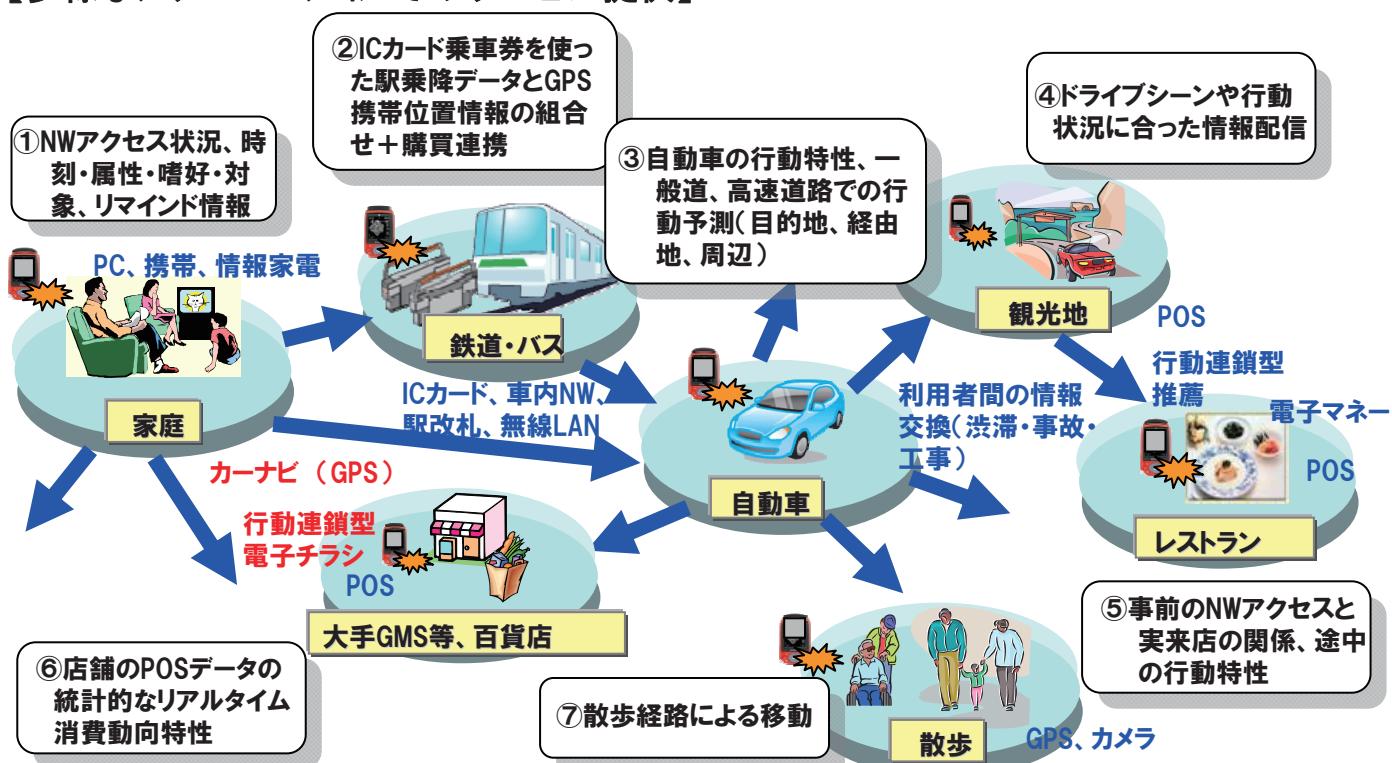
#### 4.3.1 行動情報に基づいた最適な情報やサービスの提供

- 個人の行動情報を収集し、匿名化等によってプライバシーを保護しつつ情報開示を行うことで、個人の嗜好やTPOに適した情報やサービスの提供が可能になる
  - GPS付き携帯電話、ICカード乗車券、カーナビ等、様々な媒体から行動情報を収集するとともに、多様なシチュエーションにおいて最適な情報やサービスの提供を実現できる

## 【プライバシーを保護した行動情報活用】



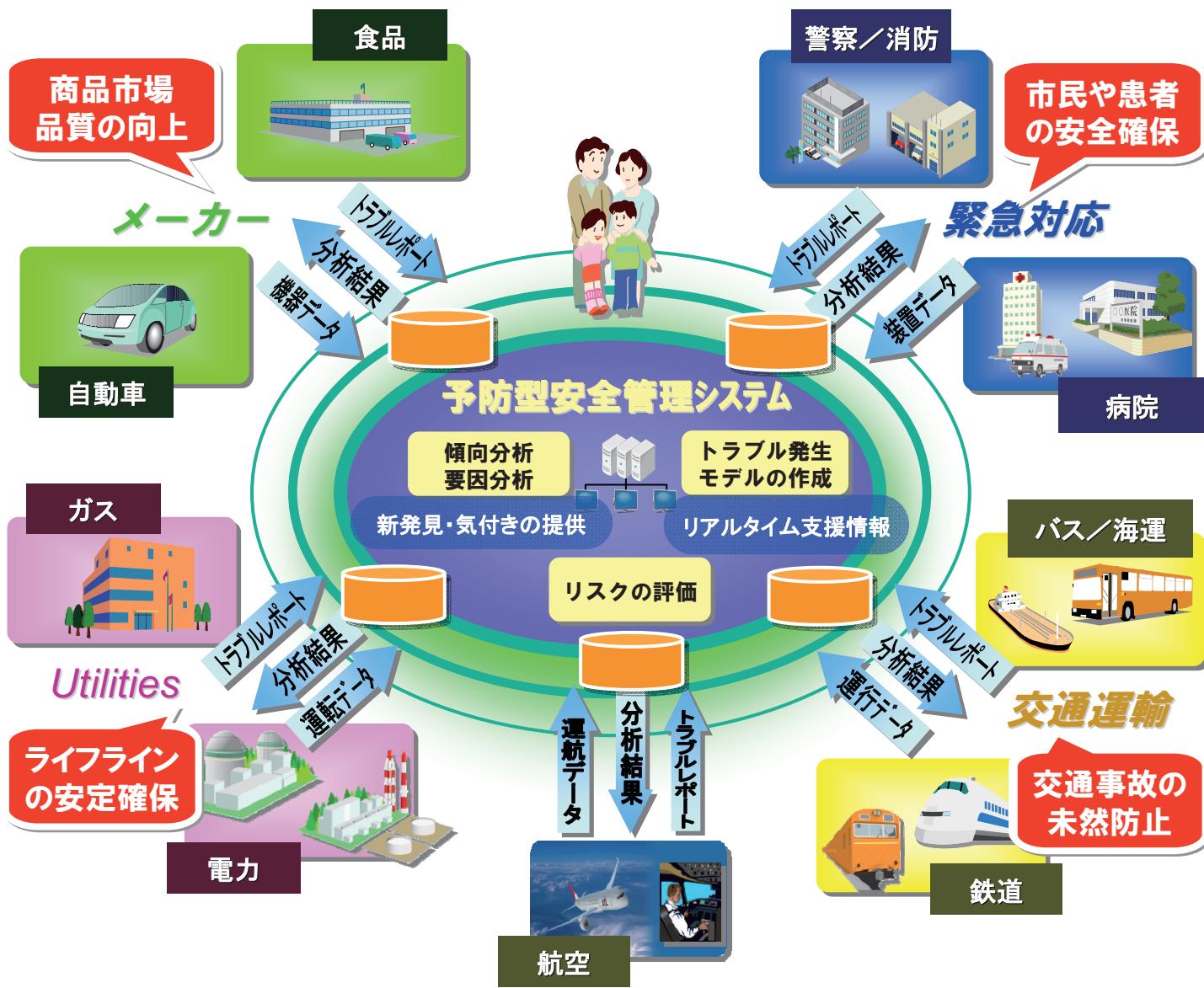
## 【多様なシチュエーションでのサービス提供】



## 4.3.2 高度な予防型安全管理システムの実現

- 企業内の大量・多様なアーカイブ情報やリアルタイム情報をデジタル融合し、高度に分析することで、より精度の高い Proactive Safety Management(予防型安全管理)を可能とする
- 陸・海・空の交通網や電力・ガスなどのライフラインだけでなく、警察、病院、メーカーなど、幅広い各業界の予防型安全管理を適用とすることで、消費者が安心して生活できる社会を実現する

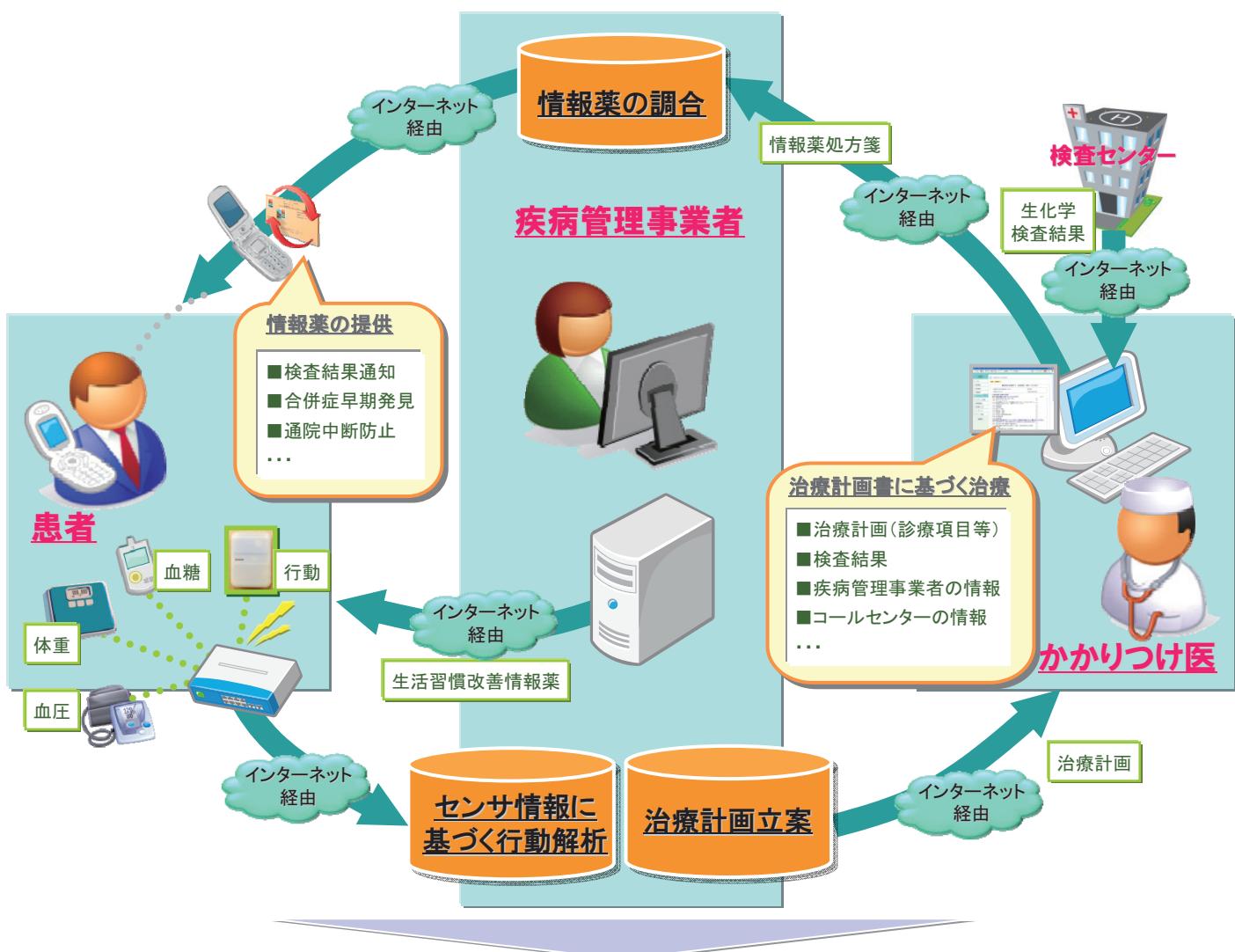
### 安心・安全な社会の実現



### 4.3.3 生活習慣に関する情報等に基づいた「情報薬」の提供

- 加速度センサ、心拍センサ、体重計、血圧計等を活用し、患者に負担を掛けない形での情報収集と行動解析を可能とし、医師・患者双方に対して適切なタイミングで適切な情報を提供することで、治療効果の向上を図る
- 専門化・高度化したガイドラインに基づく治療計画を立案し、医師に提示することで、非専門医であってもガイドラインに準拠した的確な治療を行うことが可能である
- 「情報薬」に基づく疾病のマネジメントにより重篤化を抑制することで、患者の生活の質(QOL)の維持・向上を図り、医療費の削減に貢献する

#### 【患者の生活習慣改善等に効果のある「情報薬」の提供】



疾病の重篤化抑制により、患者の生活の質向上、医療費削減に貢献

## 4.3.4 ブログランキングサービス

重要かつ信憑性の高いブログ記事に、より簡単に、より早く到達する

### 技術の概要

- 情報発信者(ブロガー)が発信した過去の記事を分析することで、この発信者のセンチメント分析を行うと共に、ある特定分野に関する熟知度を算出する
- 従来のブログ検索エンジンでは、記事のランキングを精度良く行うことができず、記事の新着順や、サイトの人気度に基づいたランキングを行うに留まっていた
- それに対し、熟知度の高い情報発信者の記事は重要性および信憑性が高いという考えに基づき、情報発信者の熟知度に基づくブログランキングシステムを実装した。なお、この時、情報発信者のセンチメント情報(特性情報)を同時に提示することにより、利用者が自分なりに記事の信憑性判断を行うことが可能である

### 技術の特長

- ブロガー熟知度およびセンチメント分析技術
  - 検索キーワードに関連の深い熟知ブロガーグループを表示することが可能である
  - ブログ記事のランキングを提供するとともに、各ブロガーの特性情報を提示することが可能である
  - 熟知グループの感情とブロガー全体の感情を比較することが可能である
  - 上記により、ユーザが閲覧するブログ記事の信憑性判断を支援することが可能である

**検索キーワード入力**

**関連熟知グループ一覧**

- ・検索キーワードに関連の深い熟知ブロガーグループを表示
- ・ユーザが各グループを選択することで、異なる視点からのランキングを表示

**プロガー全体の感情および特徴語**

- ・検索キーワードが示すトピックに関して世の中のプロガーがどのような反応を示しているのかを、感情語の分析結果により表示

**検索結果(熟知度ランキング)**

- ・各プロガーの熟知度に基づいて、ブログ記事のランキングを提供すると共に、各プロガーの特性情報を提示
- ・記事を書いたプロガーの特性を提示することで、ユーザが閲覧するブログ記事の信憑性判断を支援

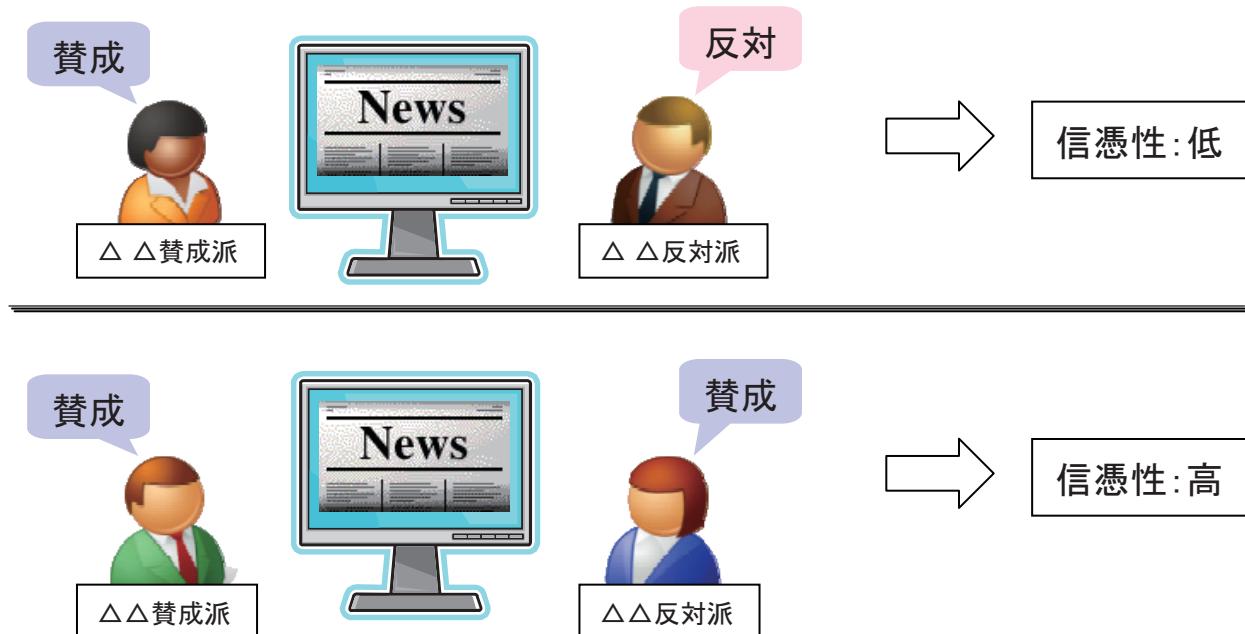
**熟知グループの感情と特徴語**

- ・検索キーワードが示すトピックに関して熟知グループがどのような反応を示しているのかを表示
- ・プロガー全体と比較することで対象熟知プロガーの特性理解を支援

プロトタイプ: [http://kizasi.jp/lab/oict\\_h20/](http://kizasi.jp/lab/oict_h20/)

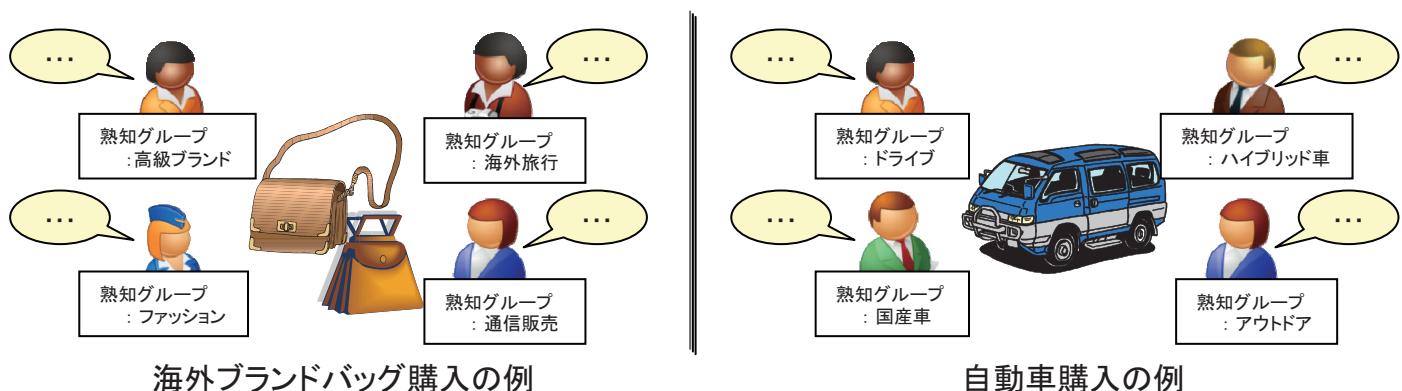
## 期待される応用分野(その1)

- Webページ等に掲載された情報(メディアの主張など)について、その関連分野に熟知しているプロガー集合のブログ記事を時系列的に解析することで、その情報(主張)の信憑性スコアを提示するサービスを実現する
  - 普段の感情値が相対するようなプロガー集合間においても、対象の情報(主張)に関して共に肯定的であれば、その報道内容等の信憑性スコアを高く算出する
  - バイアスのかかった情報などについて、その信憑性を客観的に判断することが可能である



## 期待される応用分野(その2)

- ユーザがショッピング等を行う際に、対象商品に関する幾つかの分野に熟知したプロガーの信頼性の高いクチコミ記事を提供することで安心できるショッピングサービスを実現する
  - 一つの商品に対して複数の視点からみた信頼性の高いクチコミを参考にすることが可能である



## 仕様

項目	内容
入出力のデータ形式	テキストデータ
利用形態	Webブラウザからのインターネットアクセス
稼働環境	Webブラウザを有する、Windows、Mac OSおよびUNIXマシン(要:インターネット環境)

## 4.3.5 センシングWeb

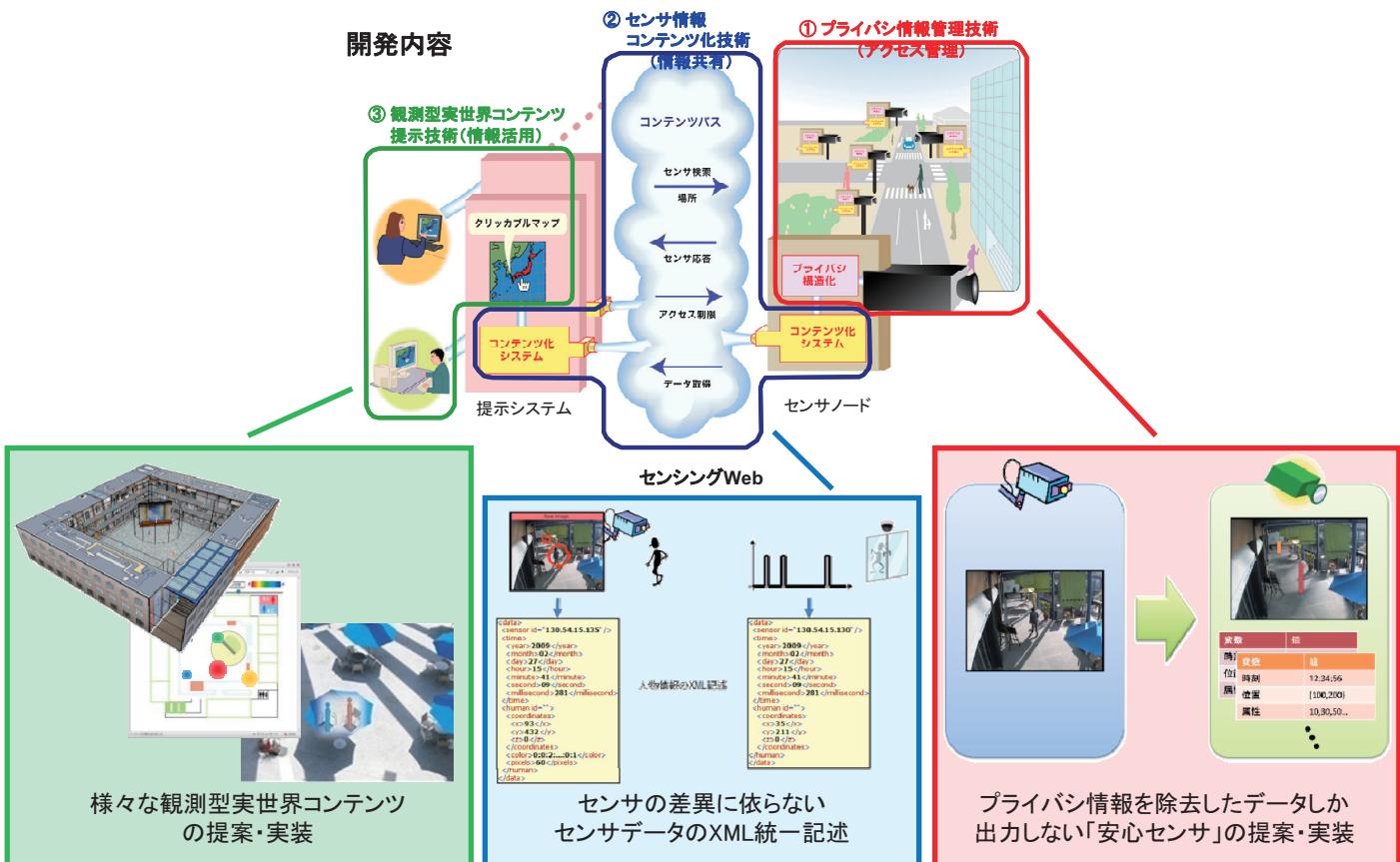
防犯カメラ等のデータを、プライバシを保護しつつ、誰でも閲覧・利用できる技術

### 技術の概要

社会の様々な場所に設置されたセンサから得られる実世界の観測情報を、被観測者のプライバシに配慮しつつ、Webのように誰もが自由に利用できるための仕組みを実現する

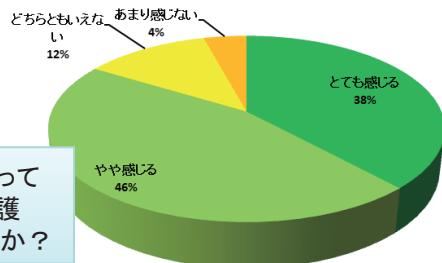
- センサ情報中の被観測者のプライバシ情報を取り除いて提供するための技術
- センサの種類や設置状況によらず多様な情報要求を受理できるための技術
- 各センサから提供される情報を統合してわかりやすく提示するための技術

### 技術の特長

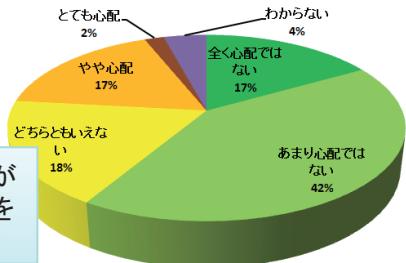


### 実証実験でのアンケート調査

「変身カメラ」によって  
プライバシー保護  
されていると感じるか？

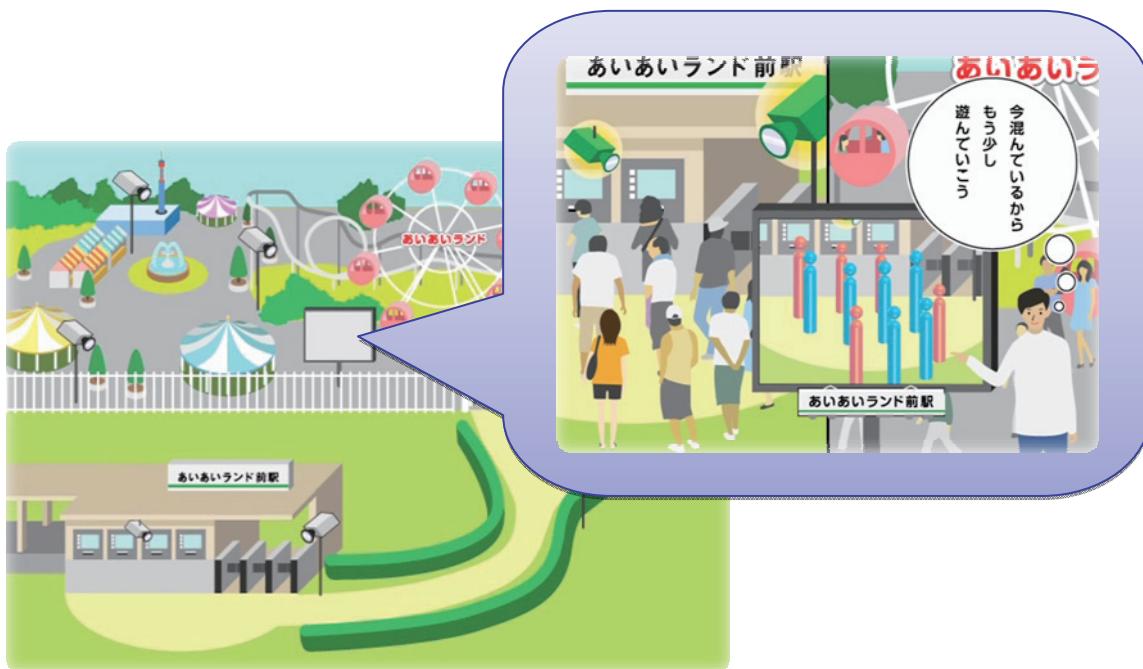


「変身カメラ」技術が  
社会に広がることを  
心配に感じるか？



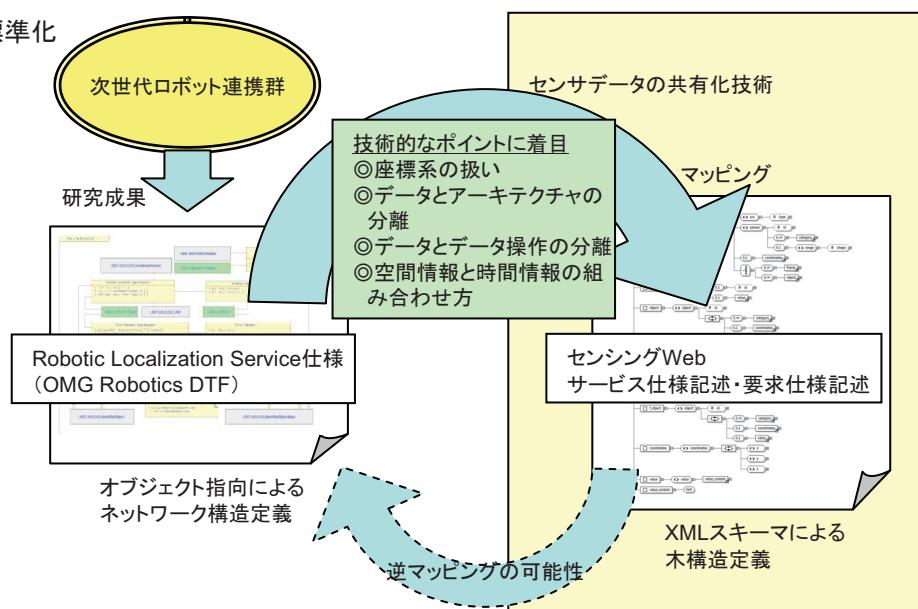
## 期待される応用分野(その1)

- 遠隔地のリアルタイムの混雑度や天候等の情報を誰でも簡単に取得できる「あいあいネット」を実現する



## 期待される応用分野(その2)

- センサデータ記述形式の標準化



## 仕様

項目	内容
センサ設置環境	新風館(京都府京都市中京区烏丸通姉小路下る場之町586-2)
センサの種類	カメラ、赤外線センサ、RFIDタグ、気象センサ
センサデータ形式	XML形式
アクセス管理	APIキー利用によるアクセス管理
URL	<a href="http://mm.media.kyoto-u.ac.jp/sweb/">http://mm.media.kyoto-u.ac.jp/sweb/</a>



## 5. 連携施策群の情報発信活動紹介

## 5. 連携施策群の情報発信活動紹介

### 概要

情報の巨大集積化と利活用基盤技術開発連携群における情報発信を目的に、対外的な活動として、シンポジウムをはじめ国内、国際学会、展示会、講演会等をこれまでに進めてきた。ここでは、平成 19 年度から平成 21 年度までに、本連携群が行ってきたシンポジウムを中心に紹介する。

### 1. シンポジウム

#### a. 平成 19 年度

##### (1) 名称

「情報の巨大集積化と利活用基盤技術開発連携群の活動」－情報爆発時代に果たす日本の役割と連携強化－ シンポジウム

##### (2) 目的

連携施策群の平成 19 年度施策の事業活動および成果について、研究者ならびに一般の方々に広く知っていただくとともに、情報爆発時代に果たす日本の役割と連携強化の方向性に関して議論し、今後の研究を進めるにあたり関連関係者の協力体制の強化および新しい取り組みを誘発する機会の拡大に寄与することを目的とする。

##### (3) プログラム

・日時：平成 20 年 1 月 30 日（水）13：00～17：30

・場所：東京ステーションコンファレンス

・内容：

開会の辞

　　奥村 直樹（総合科学技術会議議員）

##### I 基調講演

　　「情報爆発時代における日本の役割」

　　喜連川 優（東京大学）

##### II 情報の巨大集積化と利活用基盤技術開発連携群の活動紹介

　　「情報の巨大集積化と利活用基盤技術開発連携群の活動報告」

　　西尾 章治郎（主監）

### III 情報の巨大集積化と利活用基盤技術開発連携群の活動－情報爆発時代に果たす日本の役割と連携強化－ シンポジウム

・各省施策の報告（総務省、経産省、文科省）

・平成 19 年度科学技術振興調整費採択課題の研究報告

　　「センサ情報の社会利用のためのコンテンツ化」 美濃 導彦（京都大学）

・今後の取り組みに向けたパネルディスカッション

　　モデレータ：安田 浩（東京電機大学）

　　パネリスト：長尾 真（国立国会図書館）、

　　喜連川 優（東京大学）、土井 美和子（東芝）、美濃 導彦（京都大学）、

　　八尋 俊英（経済産業省）

閉会の辞 松山 隆（京都大学）

##### (4) 実施結果

###### (a) 参加者数及び内訳

参加者数：264 名（内訳：企業 165 名、大学・法人・一般 53 名、官公庁関係 46 名）

###### (b) 総括

本シンポジウム開催により、一般の方々にも広く本連携群の活動内容を理解していただいた（アンケート結果で約 8 割が「参考になった、大変参考になった」との回答）。また、本施策担当者ならびに関連する研究者間のコミュニケーションも促進することができ、今後のさらなる連携活動に資することが出来た。

### (5) シンポジウム模様



#### b. 平成 20 年度

##### (1) 名称

「情報の巨大集積化と利活用基盤技術開発連携群の活動」－情報爆発時代におけるソリューションと連携強化－ シンポジウム

##### (2) 目的

総合科学技術会議 科学技術連携施策群  
「情報の巨大集積化と利活用基盤技術開発連携群」の平成 20 年度施策の事業活動および成果について、研究開発者のみならず広く一般の方々に知っていただくことを目的とする。特に、情報爆発時代における重要な諸課題に対し、我が国がいかに世界をリードできるか等の観点からの解決策を議論する。本シンポジウムにより、今後の研究開発を進めるにあたり関連諸機関の連携体制の一層の強化を図り、国際競争力のさらなる向上を目指す。

### (3) プログラム

・日時：平成 21 年 2 月 4 日（水）11:00～

17:40

・場所：全国社会福祉協議会灘尾ホール

・内容：

#### 第一部 デモンストレーション

「情報の巨大集積化と利活用基盤技術開発連携群」の平成 20 年度施策の主な成果についてデモ展示を行う。具体的には、総合科学技術会議における貢献軸の観点である「安全・安心」、「社会」、「産業」、「科学」に対する連携群としてのソリューションのデモ展示を行う。これらの展示により、一般の方々にも連携群の活動内容の理解を深めていただく。

#### 第二部 シンポジウム

##### 開会の辞

奥村 直樹（総合科学技術会議議員）

##### I 基調講演

「進化する ICT 社会へのチャレンジ」

宇治 則孝（NTT）

##### II 情報の巨大集積化と利活用基盤技術開発連携群の活動報告

西尾 章治郎（主監）

##### III 情報利活用を発展させる戦略と今後の展望に関するシンポジウム

##### 【技術セッション】

モデレータ：東倉 洋一（NII）

プレゼンタ：

(a) 情報大航海プロジェクト

喜連川 優（東京大学）

(b) 情報信憑性検証技術

松山 隆司（京都大学）

(c) 超高性能 DB 基盤ソフトウェア

喜連川 優（東京大学）

河村 信男（日立製作所）

(d) センサ情報の社会利用のための

コンテンツ化

美濃 導彦（京都大学）

## 【総括セッション】

### パネルディスカッション

モデレータ：安田 浩（東京電機大学）

パネリスト：荒川 薫（明治大学）

岩野 和生（日本IBM）

坂内 正夫（国立情報学研究所）

村上 輝康（野村総合研究所）

八尋 俊英（経済産業省）

閉会の辞 大江田 憲治（内閣府）

### (4) 実施結果

#### (a) 参加者数及び内訳

参加者数：256名（内訳：企業162名、大学・法人・一般43名、官公庁関係51名）

(b) 総括：成果の可視化という観点から、1部デモ、2部シンポジウムの2部構成とした。アンケート結果では、デモでは約7割が、シンポジウムでは約8割が「参考になった、大変参考になった」との回答が得られ、本連携群の活動内容及び成果に関して十分に理解していただくことが出来た。

### (5) シンポジウム模様

#### (a) デモ模様



#### (b) シンポジウム模様



### c. 平成21年度シンポジウム

#### (1) 名称

「情報の巨大集積化と利活用基盤技術開発連携群の活動」－情報爆発時代におけるイノベーション創出－ シンポジウム

#### (2) 目的

連携施策群の平成21年度施策の事業活動および成果について、研究開発者のみならず広く一般の方々に知りていただくことを目的とする。特に、情報爆発時代における重要な諸課題に対し、我が国がいかに世界をリードできるか等の観点から、今後の情報通信技術の将来ビジョンを議論する。本シンポジウムにより、今後の研究開発を進めるにあたり関連諸機関の連携体制の一層の強化を図り、国際競争力のさらなる向上を目指す。

#### (3) プログラム

・日時：平成21年12月1日（火）10:30～16:30

・場所：日本科学未来館

・内容：

##### 開会の辞

奥村 直樹（総合科学技術会議議員）

##### I 基調講演

「Smarter Planet のめざす世界とクラウド・コンピューティング」

岩野 和生（日本IBM）

##### II 総合科学技術会議科学技術連携施策群 情報の巨大集積化と利活用基盤技術開発連携施策群の活動報告

「情報の巨大集積化と利活用基盤技術開発連携施策群の成果と今後の展開」

西尾 章治郎（主監）

##### III 補完的課題の成果報告（最終報告）

「センサ情報の社会利用のためのコンテナ化」

美濃 導彦（京都大学）

## IV 「情報爆発時代におけるイノベーション創出」

### ・各省における取り組み

(a) 平成 21 年度情報大航海プロジェクトの取り組みについて

東條 吉朗（経産省）

(b) 電気通信サービスにおける情報信憑性検証技術に関する研究開発

川村 一郎（総務省）

(c) 超高性能 DB 基盤ソフトウェア、Web 社会分析基盤ソフトウェアに関する研究開発

後藤 祐介（文科省）

### ・パネルディスカッション：「連携施策群の総括と今後の情報通信技術の将来ビジョン」

モデレータ：安田 浩（東京電機大学）

パネリスト：木俵 豊（NICT）

喜連川 優（東京大学）

東倉 洋一（国立情報学研究所）

松山 隆司（京都大学）

横田 治夫（東京工業大学）

鷲尾 隆（大阪大学）

岩爪 道昭（NICT）

西尾 章治郎（主監）

金谷 学（内閣府）

閉会の辞 大江田 憲治（内閣府）

### (4) 実施結果

#### (a) 参加者数及び内訳

参加者数：222 名（内訳 企業 127 名、大学・法人・一般 65 名、官公庁関係 30 名）

#### (b) 総括

シンポジウムと併設してデモ展示を実施した。アンケート結果においても、シンポジウム、デモとともに約 6 割強が「大変参考になった、参考になった」との回答が得られ、本連携群の活動内容及び成果に関し十分に理解していただくことが出来た。特に、パネルディスカッションでは、全タスクフォ

ース委員が登壇し、3 年間の活動成果の総括を行い、フロアから多くの質問、意見が寄せられ、最終年度のシンポジウムを盛況に終えることが出来た。

### (5) シンポジウム模様

#### (a) デモ模様



#### (b) シンポジウム模様



- d. 情報処理学会 一創立 50 周年記念  
 (第 72 回) 全国大会出展
- (1) デモンストレーション
- ・日時：平成 22 年 3 月 9 日、10 日
  - ・会場：東京大学 本郷キャンパス  
 御殿下記念館
  - ・内容：「情報の巨大集積化と利活用基盤技術開発」連携施策群の各施策、情報信憑性検証技術、超高性能データベース基盤ソフト、情報大航海プロジェクトおよび補完的課題（センシング Web）の成果を 2 台の大型ディスプレイを利用して動態展示する。
- (2) 特別セッション
- ・セッション名：  
 センシング Web とプライバシマネジメント  
 ー プライバシに配慮したセンサ情報の社会的利用ー
  - ・日時：3 月 10 日（水）15：30-17：50
  - ・会場：法文 1 号館 1F 22 教室
  - ・プログラム：  
 基調講演 「WEB 時代と ID・プライバシ管理」 安田浩（東京電機大学）  
 活動報告 「情報の巨大集積化と利活用基盤技術開発連携群の活動報告」  
 西尾章治郎（大阪大学）  
 講演  
 「画像情報に対するプライバシ情報処理」 美濃導彦（京都大学）  
 「音声情報に対するプライバシ情報処理」 中川聖一（豊橋技術科学大学）  
 「センサ情報に基づく実世界コンテンツ」 馬場口登（大阪大学）  
 「センサ群からの情報の構造化」 谷口倫一郎（九州大学）、  
 「センサ情報利用によるシースルービジョン」 大田友一（筑波大学）  
 「センサ情報の共有」 鯉坂恒夫（和歌山大学）

- e. 平成 21 年度 情報大航海プロジェクト  
 シンポジウム ー 情報爆発時代のイノベーション創出に向けた 3 カ年の取組と成果についてー
- ・日時：平成 22 年 3 月 8 日（月）  
 10:00～17:30 (開場 9:30)
  - ・会場：東京大学（本郷）安田講堂
  - ・プログラム：  
 10:00～11:30 説明会 平成 21 年著作権法改正のポイント 壇 俊光 氏（北尻総合法律事務所 弁護士）  
 12:30～12:35 挨拶 経済産業省  
 12:35～13:25 招待講演 First industrial and technical results from Quaero.  
 ー 欧州検索エンジンプロジェクト Quaero の成果と今後の方向性ー Pieter van der Linden 氏（Thomson Program Manager）  
 13:25～14:35 講演① 情報大航海プロジェクトの取組と成果 喜連川 優 氏（東京大学生産技術研究所 教授）  
 開発実証事業 ((株)NTT ドコモ、沖電気工業(株)、(株)キューデンインフォコム、(株)日本航空インターナショナル)  
 14:50～15:40 講演② 情報大航海プロジェクトにおける共通技術開発の取組と成果 長谷山 美紀 氏（北海道大学 教授）  
 共通技術開発企業等（チームラボ(株)、(株)データクラフト、日本電信電話(株)、パナソニック システムネットワークス(株)）  
 15:40～16:40 講演③ 情報大航海プロジェクトの制度面での取組と成果 一パソナル情報と著作権に係る制度への取組と成果ー 堀部 政男 氏（一橋大学 名誉教授）、牧野 二郎 氏（牧野総合法律事務所 弁護士）、壇 俊光 氏（北尻総合法律事務所 弁護士）、小向 太郎 氏（(株)情報通信総合研究所）、坂下 哲也 氏（(財)日

## 本情報処理開発協会

16:40～17:10 講演④ 成長戦略とポスト  
情報大航海（仮題）

### 2. ホームページによる情報発信

図1及び2に示すように、連携群におけるホームページを開設し、各府省の施策紹介、これまでのシンポジウム資料の掲載等の情報発信を行っている。また、図3に示すように、国際化の観点より英語版のホームページも並行して開設し、同様に情報発信を行っている。

The screenshot shows the homepage of the "Information Utilization Collaboration Group". The top navigation bar includes links for Home, Message, and Results from平成19年度(2007), 平成20年度(2008), and 平成21年度(2009). A large globe graphic highlights Japan. The main title is "情報の巨大集積化と利活用基盤技術開発連携群の活動" (Activities of the Large-scale Information Integration and Utilization Platform Technology Development Collaboration Group). Below the title, there are sections for "各府省の委託事業等紹介" (Introduction of entrusted projects by various ministries and agencies) and "Index". A sidebar on the left lists links for Home, Message, and Results from 2007, 2008, and 2009. At the bottom, there is a footer with a back-link to the homepage and a link to the English version.

図1 連携群ホームページ（トップページ）

The screenshot shows the English version of the homepage. The top navigation bar includes links for Home, Message, and Results from平成19年度(2007), 平成20年度(2008), and 平成21年度(2009). A portrait of Mr. Nishio is displayed with the text "連携施策群コーディネーター／主監 西尾章治郎からのメッセージ" (Message from Coordinator/Supervisor Mr. Nishio). The main content area discusses the "Large-scale Information Integration and Utilization Platform Technology Development Collaboration Group" and its activities. It mentions the use of Web technologies for information integration and utilization, and the promotion of international cooperation through English-language websites.

図2 連携群ホームページ（メッセージ）

This screenshot is identical to the one in Figure 2, showing the English version of the homepage. It features the same navigation bar, Mr. Nishio's portrait, and the main content about the collaboration group's activities. The English text provides a detailed explanation of the group's goals and the use of Web technologies for information integration and utilization.

図3 連携群ホームページ（英語版）

おわりに

内閣府 参事官（情報通信担当）  
金谷 学

本報告は、平成 19 年から内閣府総合科学技術会議科学技術連携施策群として推進してきた「情報の巨大集積化と利活用基盤技術開発」連携群の最終年度としての成果をとりまとめたものです。

日々、個人レベルでの日常生活から企業等組織レベルでの経済社会活動に至るまで、世界中で莫大な情報が生み出され蓄積されています。そして、その新たな情報 자체が次の情報を生み出す源ともなっています。

このような情報の巨大集積に関し、国民が安心して的確な情報を収集、利活用できるよう、内閣府では関係省庁である総務省、文部科学省、経済産業省の関連研究施策の密接な連携を図ってきました。しかし、連携施策を実施してきた 3 年の間にも情報の集積は益々巨大になり、これに伴う技術課題も多様で高度なものとなってきています。

本連携施策群としての活動は、平成 21 年度で一区切りをつけることとなります。今後とも情報通信関連技術の革新により、新たな情報蓄積手段や利活用手段が次々と出現していくものと思われ、我が国として情報の巨大集積化と利活用への取り組みを今後とも行っていくことが重要です。その際、情報が巨大集積化したことへの対処という面のみならず、従来では困難であった社会システム、制度の改革について研究開発面から貢献していくという姿勢も重要であると考えています。

最後に、本連携施策群のコーディネーターおよび主監を務めていただいた大阪大学理事・副学長の西尾章治郎先生、タスクフォース委員の諸先生方、連携施策を実施した総務省、文部科学省および経済産業省、そして（独）科学技術振興機構の科学技術連携施策群支援業務室の方々に深甚なる感謝の意を表する次第です。