

「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

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」
主監 西尾 章治郎

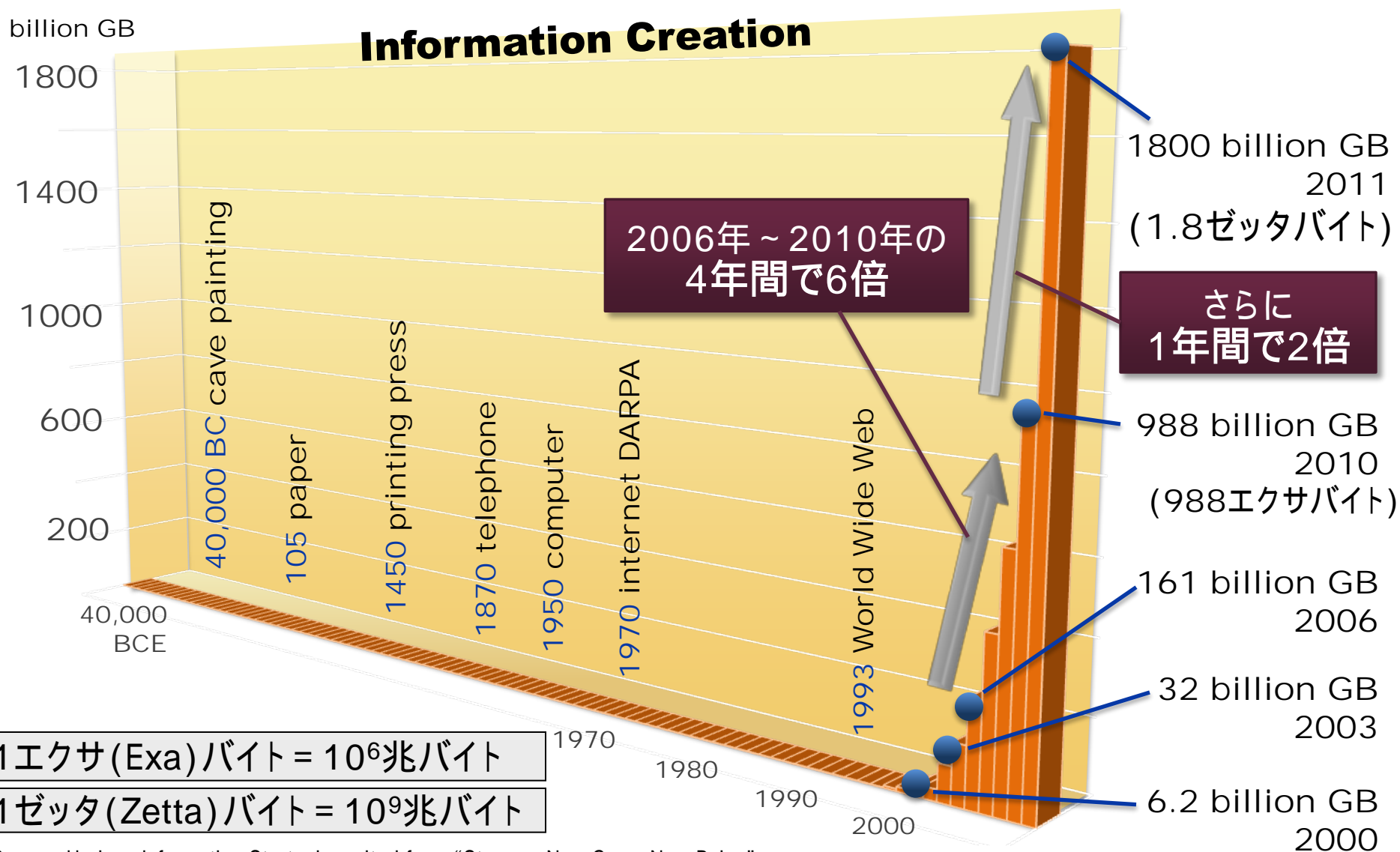
平成21年12月1日

報告内容

1. 背景・目的
2. 情報の巨大集積化と利活用基盤技術開発連携群
3. 各省施策及び補完的課題における主な成果
4. 連携群の活動における3年間の総括
5. 今後の展開

1. 背景·目的

情報爆発時代の到来

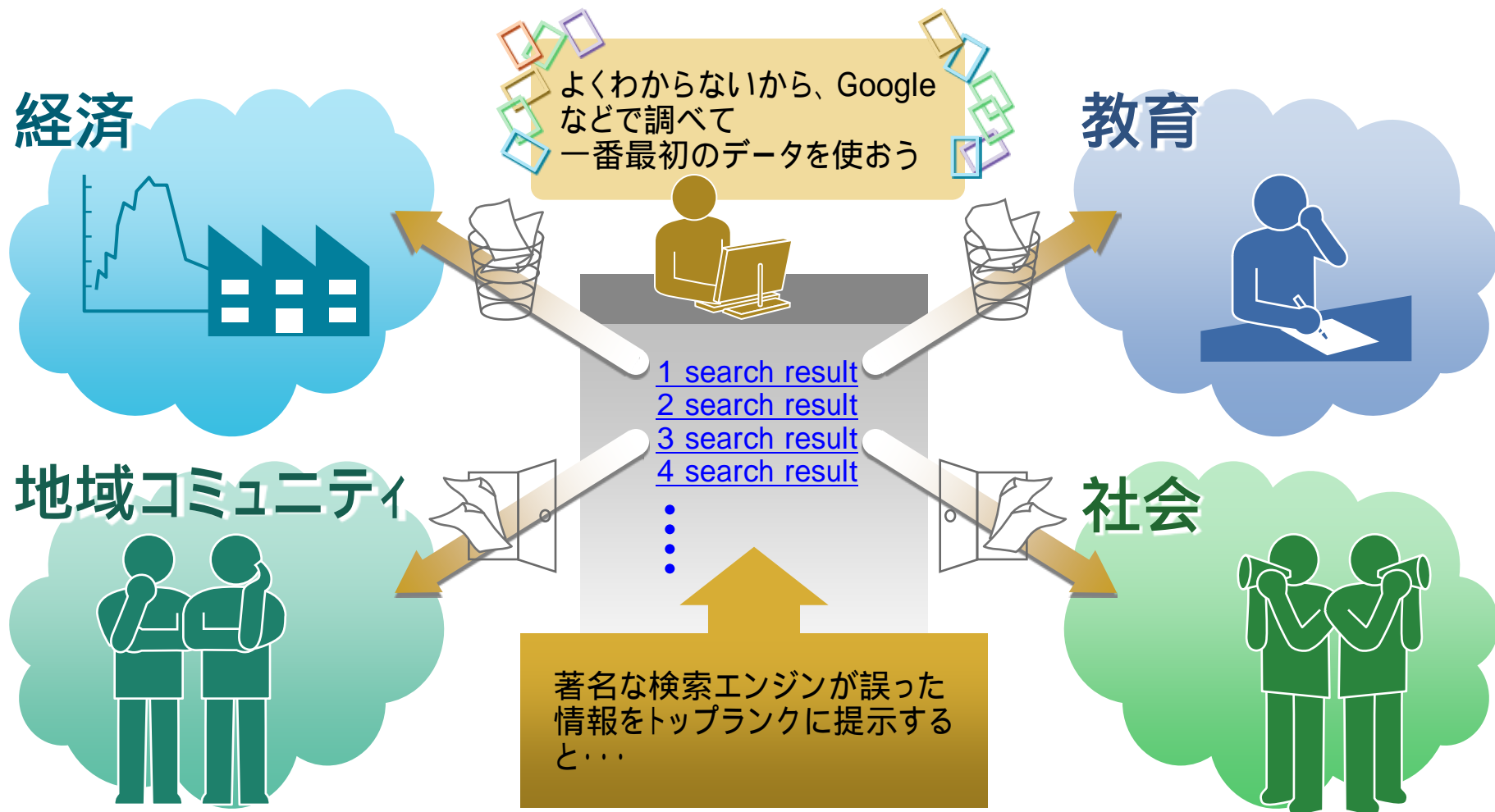


Source: Horison Information Strategies, cited from "Storage: New Game New Rules"

Source: Information Data Corporation, "The Diverse and Exploding Digital Universe", 2008

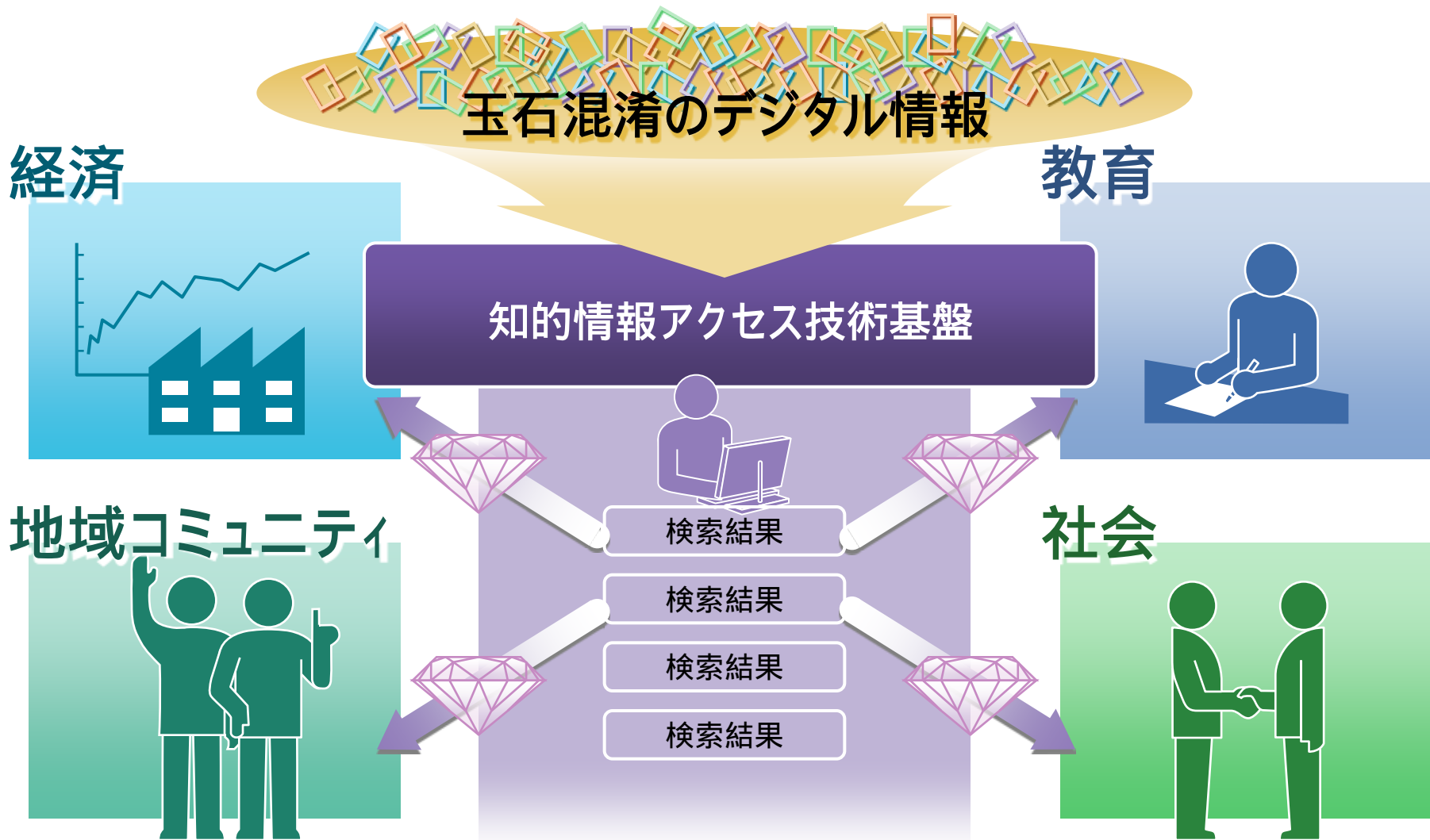
従来の検索サービスの社会的な問題の例

- 今の検索エンジンは、ユーザの観点から情報を分析して提示していない。
- その結果、ゴミ情報や有害な情報を上位ランクに提示する危険性がある。



知的情報アクセス技術基盤を活用した社会

目的：知的なサービスの活用で安全・安心な社会を実現



2. 情報の巨大集積化と 利活用基盤技術開発連携群

科学技術連携施策群の推進体制

総合科学技術会議（本会議）

基本政策推進専門調査会

内閣府

分野別推進戦略総合PT

科学技術連携施策群（6プログラム）

科学技術政策
担当大臣

ライフサイエンスPT

n 食料・生物生産研究 n 臨床研究・臨床への橋渡し研究

情報通信PT

n 情報の巨大集積化と利活用基盤技術開発

環境PT

n 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

n 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

n テロ対策のための研究開発

政策統括官

事務局

「情報の巨大集積化と利活用基盤技術開発」連携群 平成19年度より4プロジェクトで開始

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。

各省連携戦略プロジェクト

文科省

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
(戦略重点)

経済産業省

情報大航海プロジェクト
(戦略重点)

総務省

電気通信サービスにおける
情報信憑性検証技術等に
関する研究開発
(戦略重点)

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
(平成19年度採択課題)

平成19年度

5年計画

平成23年度

平成19年度

3年計画

平成21年度

平成19年度

4年計画

平成22年度

平成19年度

3年計画

平成21年度

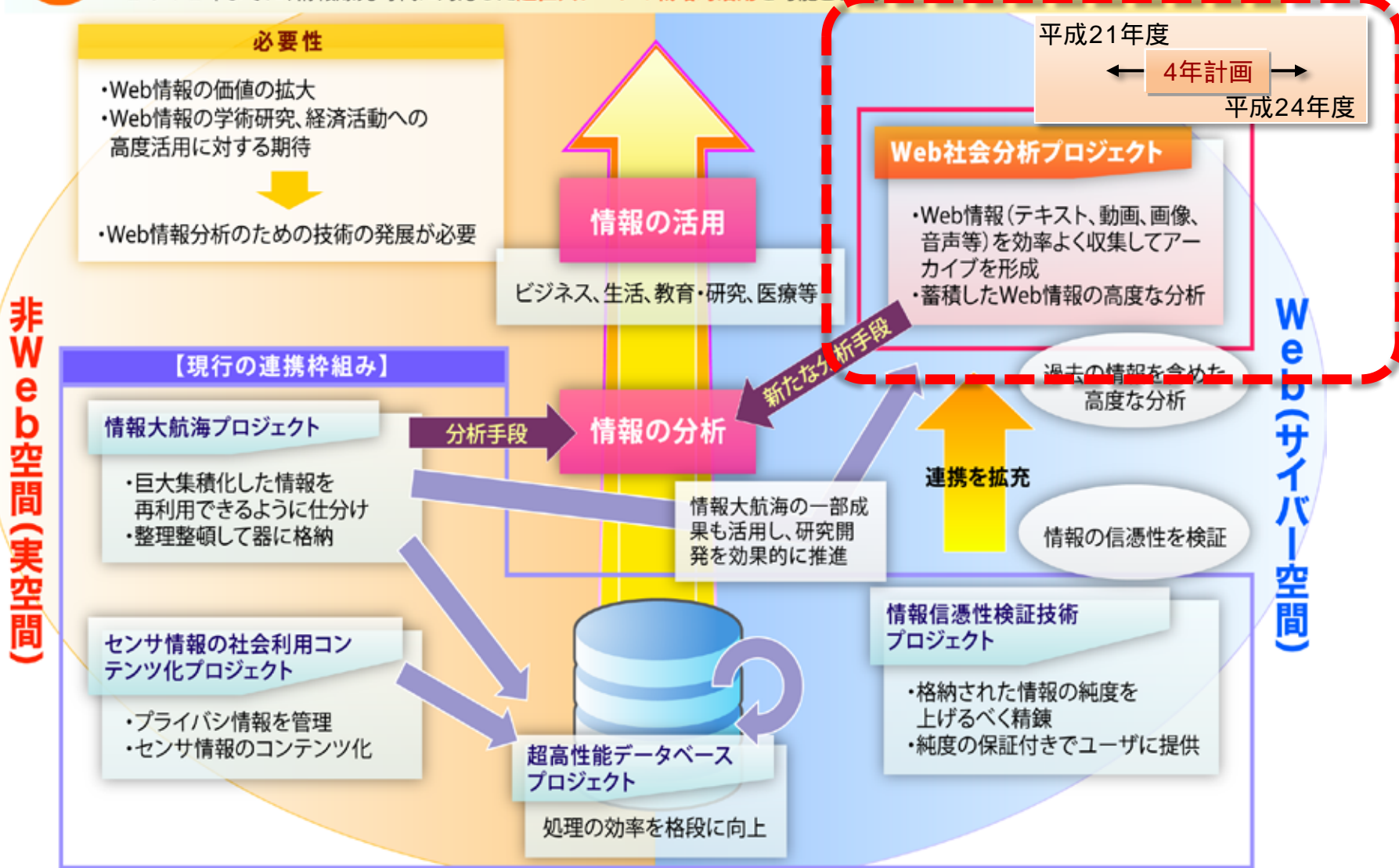
「情報の巨大集積化と利活用基盤技術開発」連携群の施策拡充（平成21年度より5プロジェクトに）

目標

Web上及び非Web上にある大量かつ多様な情報を、個人が簡便、的確、かつ安心して収集、分析できる

次世代の情報検索・情報解析技術基盤を構築する。

2012年までに、情報爆発時代に対応した超巨大データの戦略的活用を可能とする。



3. 各省施策及び補完的課題における 主な成果

情報大航海プロジェクト

経済産業省

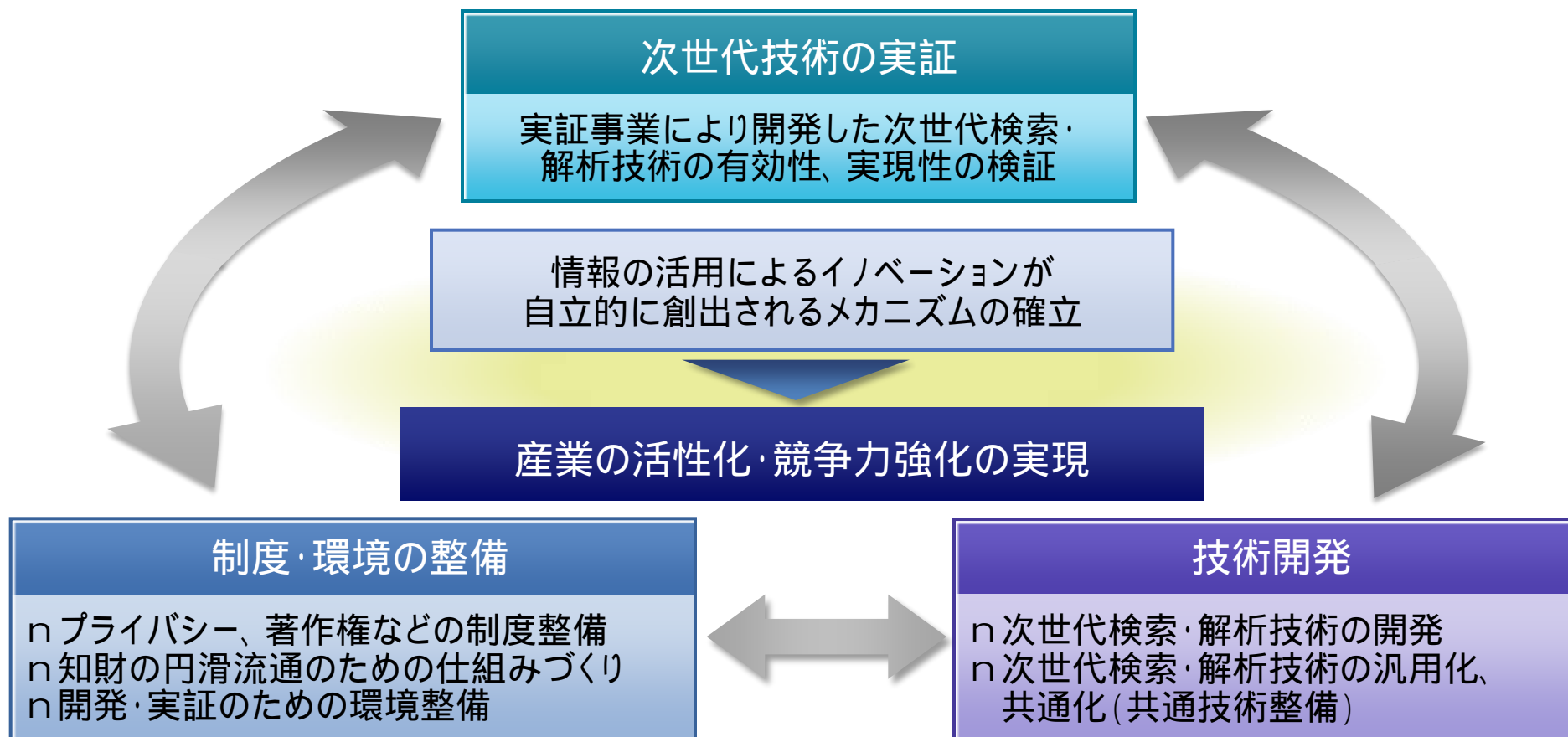
<http://www.igvpj.jp/index/>

http://www.igvpj.jp/index_en/

情報大航海プロジェクト

目標

次世代技術に係る事業の成功をトリガーに、「**制度・環境**」の更なる成熟、「**技術**」の更なる発展を喚起し、より先進的な事業の創出へつながる**イノベーション創出メカニズム**を確立し、**産業の活性化、競争力強化**を実現する。



主な成果：次世代技術の実証

次世代技術の実証

以下に示す三つの方向性から延べ22の実証事業を実施し、実現性の高いサービスであるとの知見を得た。

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

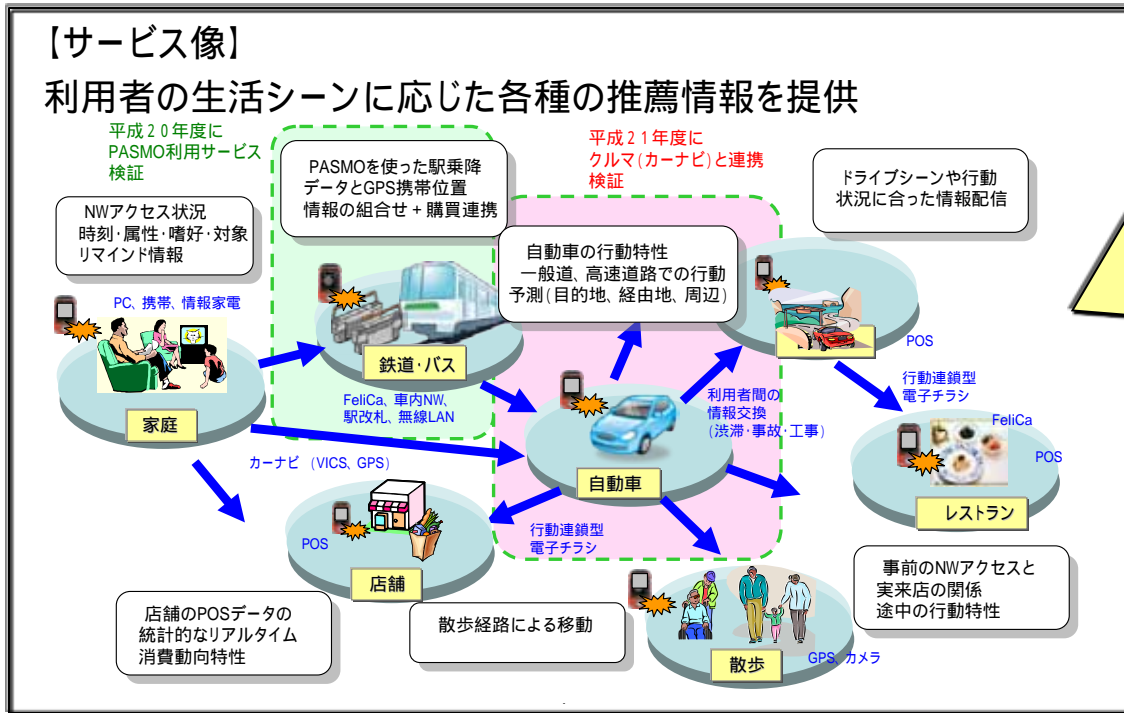
次世代検索・解析技術を活用して、事故や疾患の可能性を事前に察知し、対応する仕組みを構築することによって、様々なリスクを回避・軽減することが可能となる。

次世代技術の実証の具体例

マイ・ライフ・アシストサービス: (株)NTTドコモ

ケータイなどから収集される普段の行動情報に基づき、新しい気付きや発見につながる情報を発信するサービス“マイ・ライフ・アシストサービス”を実施し、サービスの有益性を検証

今後、本サービスで得たノウハウを“iコンシェルサービス”に展開



【サービスイメージ】

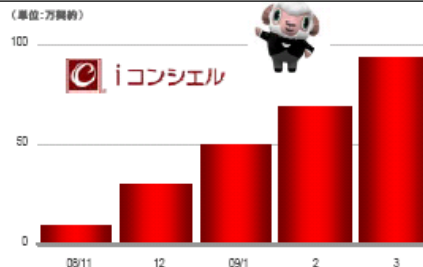
ケータイを覗いたときに、自分の行動にマッチした情報がさりげなく表示

行動履歴などから行動を推測して気付きを提案



iコンシェル

240万契約突破
 (2009年10月14日現在)



09年度目標: **380万契約**

主な成果：共通技術

共通技術

産官学で構成したタスクフォースと連携して重点化項目、開発内容を策定し、汎用性・共通性が高い次世代検索・解析技術として、サービスに直結した**実用技術**と、それを補完する**先端の基盤技術**を開発（50技術を超える技術を開発）

サービス 共通技術

開発実証企業が開発するサービスに直結した実用技術であり、汎用性、共通性が高く、次世代検索・解析技術として抽出・共通化されるべき技術（産・学のタスクフォースとベンチマーク調査により重点化項目を決定）

基盤 共通技術

開発実証企業では開発しないが、将来の社会生活基盤において必要な技術であり、サービス共通技術を補完する最先端の研究に基づく次世代検索・解析技術（産・官・学のタスクフォースと審議会が開発内容を決定）

マイ・ライフ・アシストサービス

交通系非接触ICカードを活用した連携型サービス

プロフィール
パスポート

ラダリング型
検索サービス

多言語対応動画
アプリケーション
プラットフォーム

Viewサーチ
北海道

サグールTV

時空間情報
マイニングサービス

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

すこやかライフ
サポートサービス

新総合安全運航
支援システム

メガリサーチ

「ここなら」
コミュニケーション
サービス

異業種連携ID
による流通サービス

地域活性化を支える
空間サービス
（ぷらっととPlatform）

抽出した 共通技術



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

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

情報信憑性検証技術

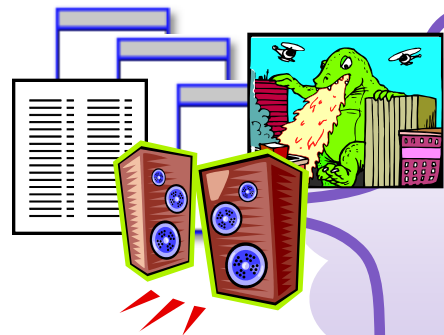
情報信憑性検証技術

ネットワーク上に流通する玉石混淆な情報を分析・提示することにより、情報利用者が個々の判断基準でWebコンテンツを容易に利活用可能となる技術

ネットワーク上の玉石混淆な情報

Webコンテンツを構成する
テキスト、画像、音声、映像情報

データ解析的アプローチによる、テキスト、画像、音声、映像情報の分析



課題ア Webコンテンツの分析技術

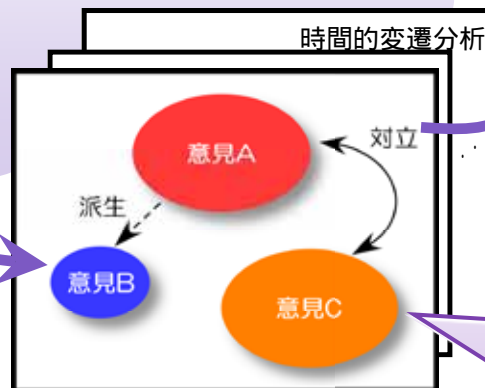
- n 画像・音声・映像情報とそれを説明するテキストの不整合度を検証するクロスメディア分析技術
- n テキストの表層的特徴(文体、語彙等)や発信者の発信履歴等の分析技術



情報利用者がWebコンテンツの信憑性を判断する上での手がかりとなる情報を提供

テキスト情報の意味内容解析を含む自然言語処理による深い解析

知識ベース



課題イ 意味内容の時系列分析技術

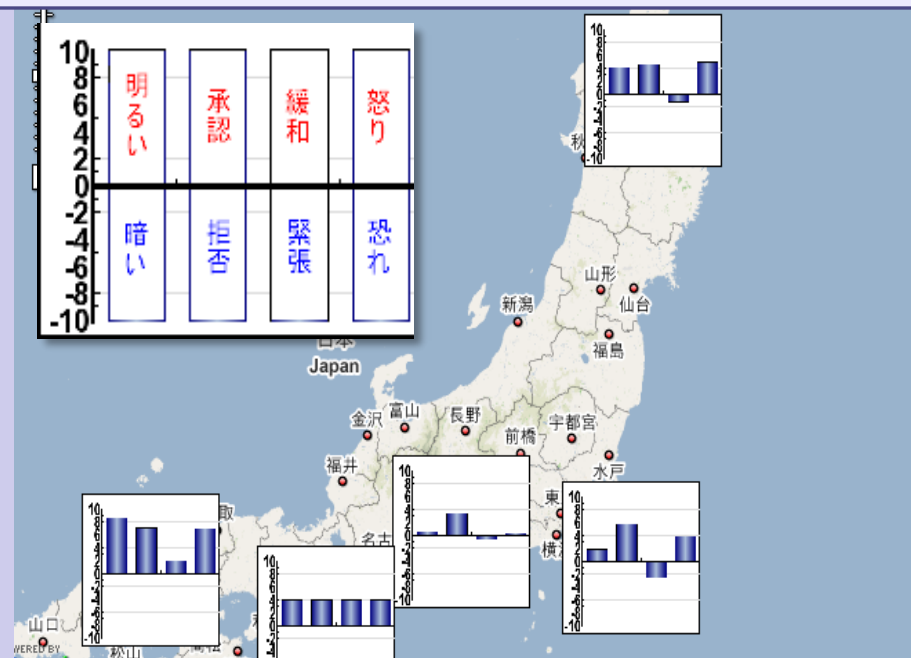
- n インターネット上から利用者が着目する領域・観点に関する意見を抽出、意見間の論理的関係を分析する技術
- n 意見の時間的変遷等を分析する技術

主な成果:「Webコンテンツの分析技術」

発信履歴情報からの発信者熟知度・センチメント分析に基づく発信者分析エンジンの開発

- 1 発信者がどのような感情（センチメント）で情報を発信しているのか？
 - 1 発信している分野にどの程度熟知しているか？
- それらを分析することで、話題に対する新聞社サイトや各国、さらにブロガーなど発信者の信憑性を計る技術の開発

1. 感情表現語辞書の構築・改良
2. 各記事のセンチメント（感情）を抽出
3. 話題ごと、発信者ごと、組織ごとのセンチメントの検出
4. 話題・発信者（個人や組織・団体）・記事などのセンチメントの相違を分析
5. 相違の提示による信憑性判断支援を実現



地方ごとの話題に対するセンチメントを分析
例) 話題:「巨人」

主な成果:「意味内容の時系列分析技術」

情報信頼性分析

n目的: 広い視野から偏りのない判断を支援
 n手段: 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



朝バナナで楽やせ!
 超簡単! その驚きの効果...

朝バナナで痩せられる!

その情報、信じてよいか?

論理空間に位置づける
 他のページはどうしているか?

反対意見は?

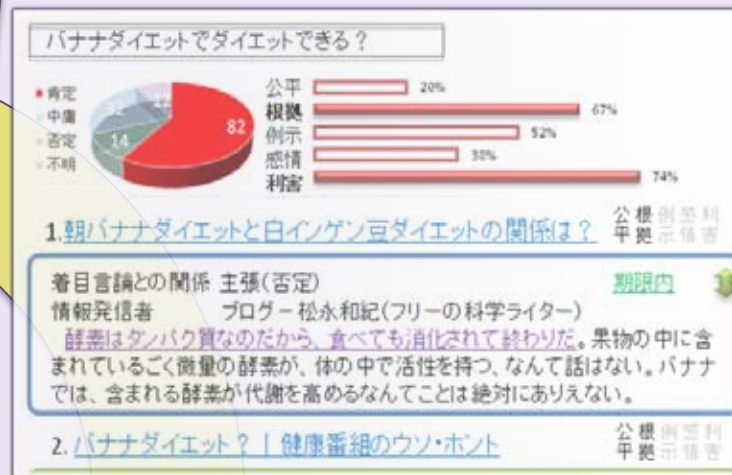
根拠は何か?

時間軸に位置づける
 どんな経緯で語られたのか?

何がきっかけか?

今でも有効か?

分析結果を統合的に出力



言論空間を再構成する
 偏りのない情報を分かりやすく

どこに注目すべきか?

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

<http://cif.iis.u-tokyo.ac.jp/OoODE/#>

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

目的

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等、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「超高性能データベース基盤ソフトウェア」を開発する。

課題

超巨大情報の戦略的活用は、国民の安心・安全を支える社会基盤の確保、並びに多様な新規産業の創出などを可能とする国力の源泉となるキーテクノロジー
情報爆発時代の超巨大情報の戦略的活用には**既存データベース基盤技術の延長線では限界**

実行原理の革新による性能ブレイクスルーが必要

研究開発内容

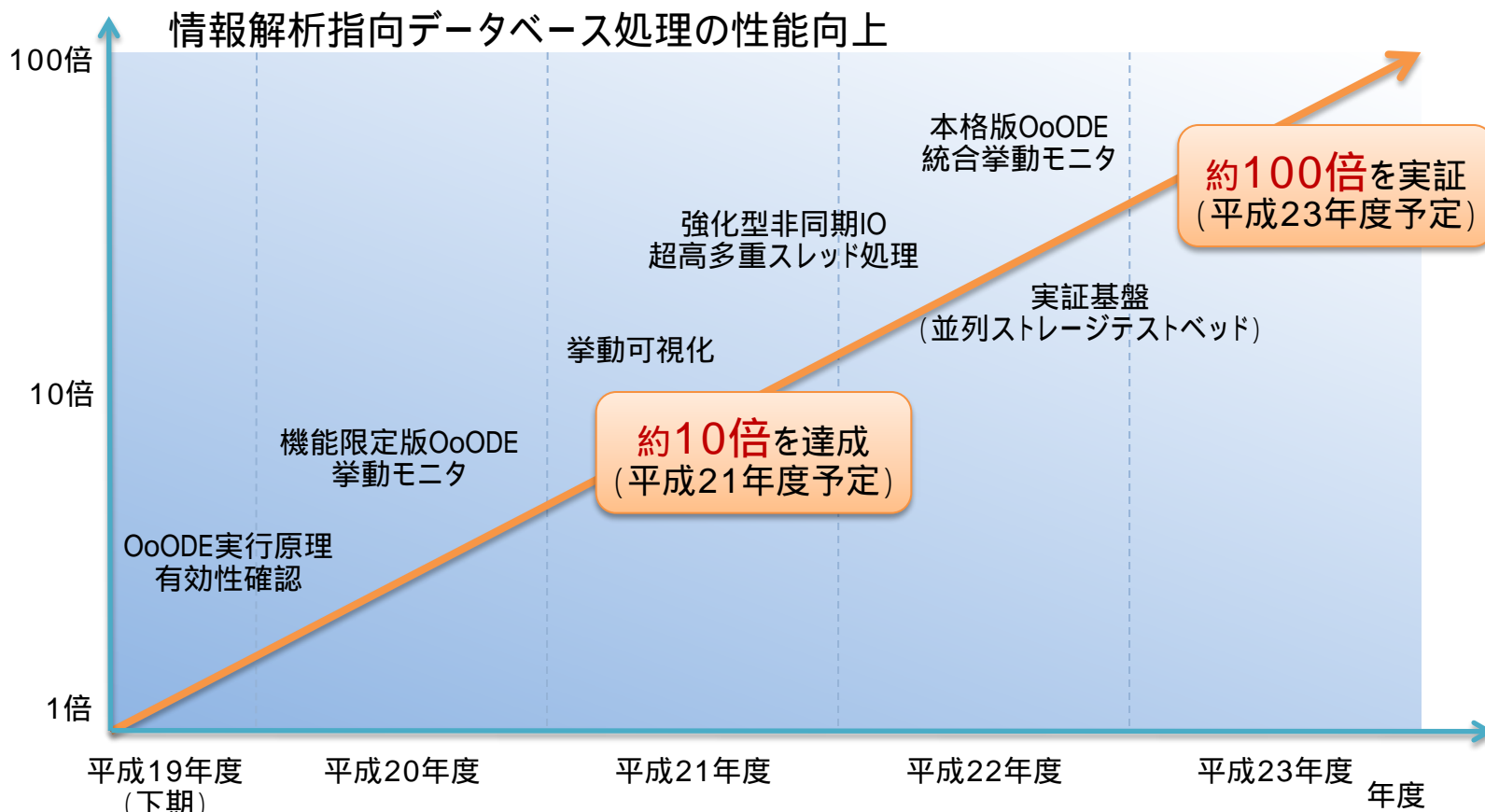
n非順序 (アウトオブオーダー) 型実行原理に基づく超高性能データベースエンジンを開発するため、以下の研究開発を実施する。

革新的実行原理である「非順序型データベース実行原理」の確立
上記 に基づくデータベース基盤ソフトウェアの設計・実装
実アプリケーションによる有効性の実証

成果目標

平成21年度中には、非順序型実行原理を一部のデータベース演算に適応したデータベースエンジンを開発し、情報解析指向のデータベース処理について従来の技術より**約10倍の性能向上**を達成する。

平成23年度中には、非順序型実行原理を本格的に適応したデータベースエンジンを開発し、情報解析指向のデータベース処理について従来の技術より**約100倍の性能向上**を達成する。



主な成果：実験結果

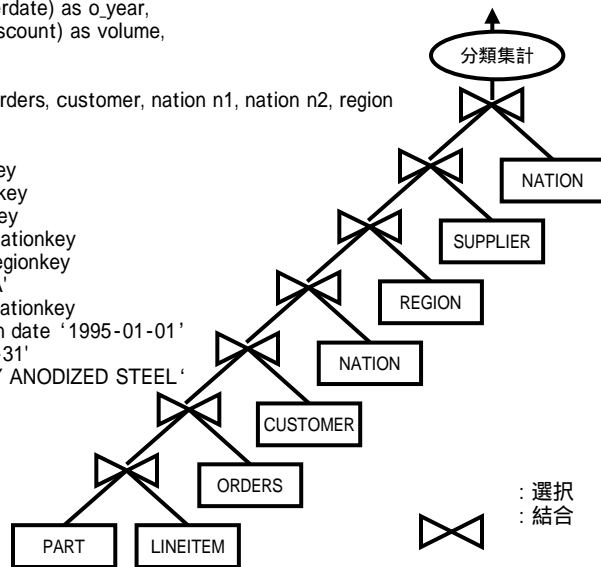
n オープンソースDBMS (MySQL) ならびに商用DBMS (HiRDB) をベースに、非順序型データベースエンジンの開発を進めてきた。このうち、商用DBMSをベースとする実装では、平成21年度内に約10倍の性能向上を実現するとの目標に対して、業界標準ベンチマークのうち複雑度の高い問合せ処理において、約20倍の性能向上を達成した。

業界標準ベンチマーク TPC-H Query 8

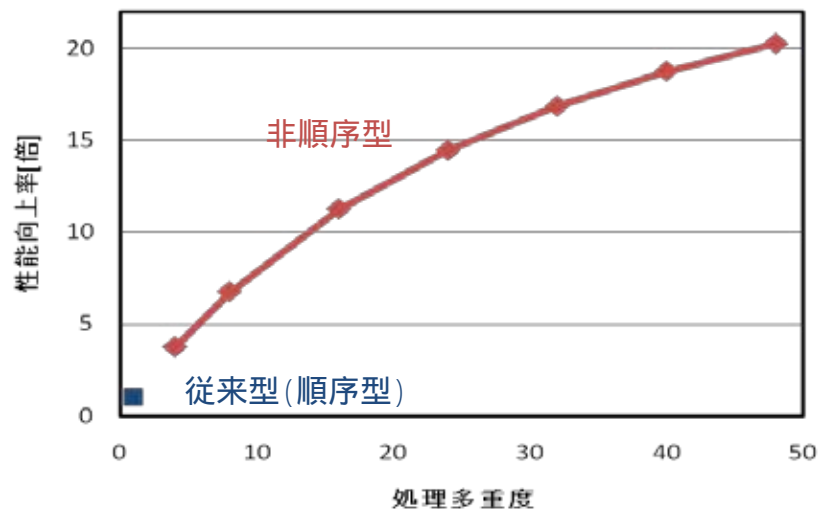
指定地域における指定商品種別の2年間の市場シェア変化を分析

```

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.n_nationkey
    and n1.n_regionkey = r_regionkey
    and r_name = 'AMERICA'
    and s_nationkey = n2.n_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;
    
```



従来型(順序型)エンジンに対する非順序型エンジンの性能向上率



プロセッサ	Intel Xeon 2.66GHz
メモリ	32GB
ストレージ	HDD (100Krpm) 20台
データセット	TPC-H SF=1000 (1TB相当)

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
(平成21年度より開始)

文部科学省

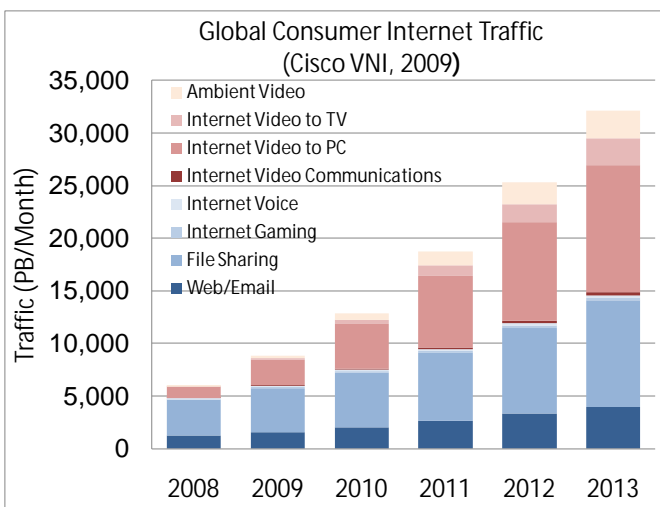
多メディアWeb解析基盤の構築及び 社会分析ソフトウェアの開発

目的

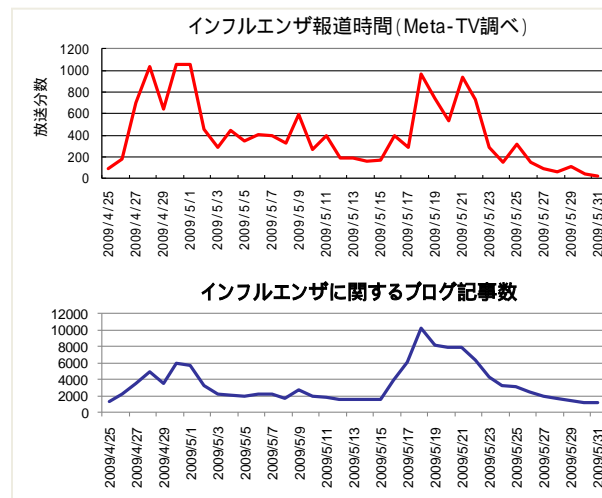
社会学、言語学、リスク管理、マーケティング等多様な社会分析ニーズに応じるために、膨大な多メディアWeb情報を収集、蓄積し多様な解析を可能とする**多メディアWeb情報解析基盤の構築**と**社会分析ソフトウェアの研究開発並びに実証**を行う。

課題

- n Web情報は人類社会の観測・調査・解析において新価値創出のために必要不可欠な情報源
 - 多メディア化が急速に進むと同時に、実世界情報と相互に及ぼし合う影響も拡大
- n 多メディアWeb情報の**収集・蓄積**、**多メディア内容解析**、**高並列計算環境上での大容量・高スループット解析基盤**、**有効な社会解析ソフトウェアの実現が必須**



動画トラフィックの増加



実社会との密接な関連

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

研究開発内容

四つのサブテーマに基づき、多メディアWebリーズニング解析を軸として、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行う。



Web情報

多メディアWeb基盤技術

索引更新処理の高速化

スケジューリング技術

多メディアWeb解析要素技術

- n キーワード抽出技術
- n リンケージ技術
- n 多次元解析高速化技術
- n 可視化技術



放送映像情報

多メディア統合処理

画像映像解析

リーズニング解析

言及の時系列解析
自然言語処理

リンク構造の時系列分析

多メディアWeb解析の実証評価

活動・書籍等

社会起業の広がり

ブログ記事数

TV番組で紹介

「世界を救う社会起業家たち！」

センシングWeb: 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(代表機関:京都大学)

<http://www.mm.media.kyoto-u.ac.jp/sweb/index.html>

研究の背景と目的

背景

- n Webは世界的な規模で進展，膨大な知識ベースとなっている。
- n センサネットワークは目的別に閉じて広く活用されているとは言い難い。
- n センサ情報を社会利用できるコンテンツに発展させるための研究が必須。

目的

センサ情報の利活用（センシングWeb）実現のため、次の機能を研究開発する。

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

主な成果：センシングWeb技術開発

センシングWeb技術の基本機能を開発

センサ情報中の被観測者のプライバシー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



プライバシー情報フィルタリング



```
<data>
<sensor id="130.54.15.135" />
<time>
<year>2009</year>
<month>02</month>
<day>27</day>
<hour>15</hour>
<minute>41</minute>
<second>09</second>
<millisecond>281</millisecond>
</time>
```

情報共有のための記述形式



センサ情報の統合・コンテンツ提示

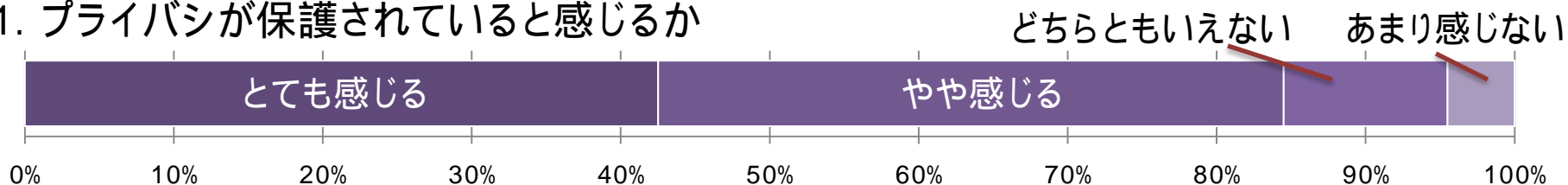


主な成果：実証実験と受容性評価

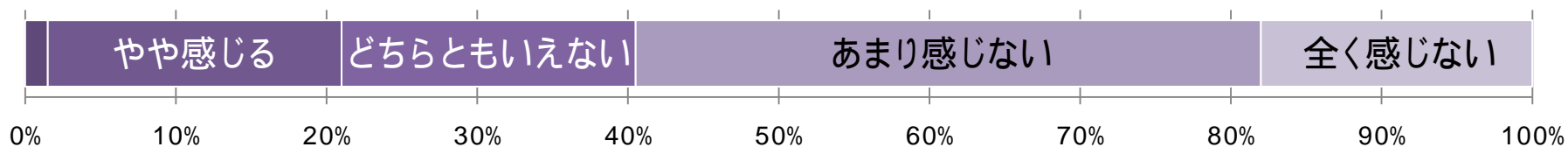
実際の公共環境での実証実験

- n 実験期間：平成21年7月～平成21年12月
- n 実験内容：常時実験により、以下を評価する。
 1. 開発した技術の実環境における性能評価
 2. 実社会におけるセンシングWebの有効性・受容性の評価

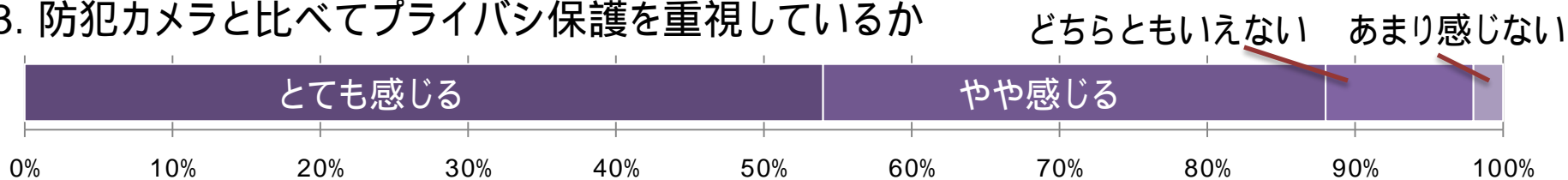
1. プライバシが保護されていると感じるか



2. この技術が社会的に広まることに心配か

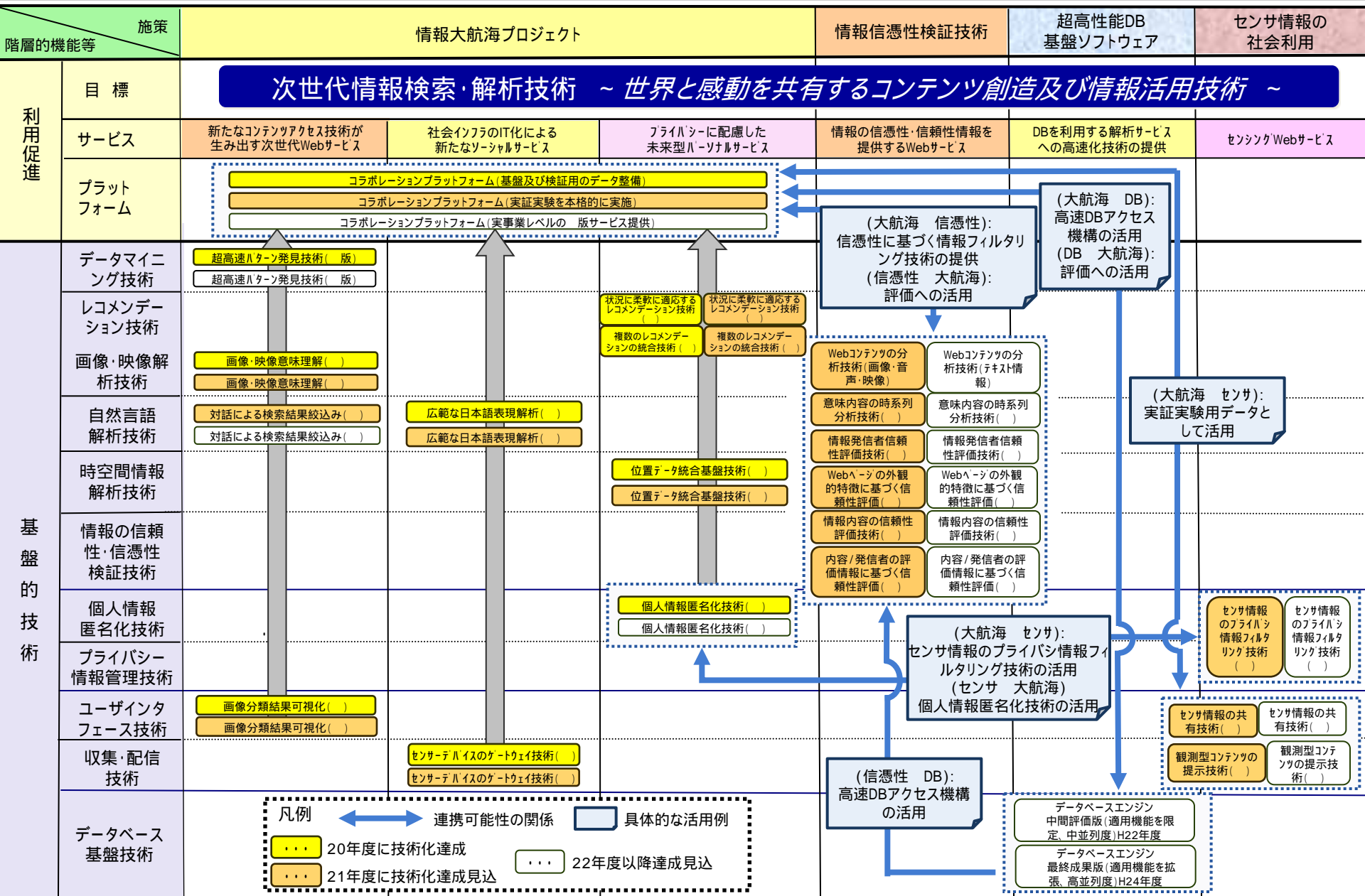


3. 防犯カメラと比べてプライバシー保護を重視しているか



各省施策及び補完的課題の連携

「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(詳細)



4. 連携群の活動における3年間の総括

「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3年間の活動の総括

(1) 施策

当初計画

1. 情報大航海
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

主な成果

- 既存施策：**着実に進展**
- 新たな施策の追加：
「**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
平成21年度より4年計画で開始

当初の目標を十分に達成！！

当初計画を大幅に上回る**高速化を実現**
桁違いの高速化につながる新たな提案へ（超高性能データベース基盤）

(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

- 著作権改正への働きかけを実施し、**著作権改正法成立**（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
コンソーシアムを設立

(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**成果の発信**（国内外）

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、ファロスからの招待講演

(1) 施策における成果の総括 (1/2)

情報大航海

n次世代技術の実証: 延べ22サービスに関して実施

n共通技術の開発: 映像・テキスト解析技術等、
50を越える基本技術を開発

n制度・環境の整備:

- 検索サービス事業が可能となる著作権法の改正
- 産学主導でパーソナル情報のサービス適用を検討する環境 (コンソーシアム) の設立

n国内講演多数
n国際会議発表
nメディア掲載: 236件

情報信憑性検証技術

nWeb情報の信憑性(画像、動画・
音声コンテンツテキスト等を開発): 実証実験の実施

nWeb発信者の信頼度分析

(言論マップ作成システム、テキスト情報の要約・整理、時系列分析の開発)

n国際ワークショップ開催
n国際会議発表: 31件
nビジネスショー等での展示

(1) 施策における成果の総括 (2/2)

革新的実行原理に基づく超高性能データベース

n 非順序型DBエンジンの開発を実施し、約20倍の性能向上を確認

(OSS DBMS、市販DBMS等実環境で検証)

○平成21年度の間目標は約10倍

n 国内シンポジウム等での発表

n 研究推進者の喜連川優教授が2009 SIGMOD Edgar F. Codd Innovations Award を受賞

センシングWeb

n センサプライバシー情報フィルタリングの実装

n センサデータの記述形式の策定、データの一般公開

n 複数の実世界観測型コンテンツの実装

n 商業施設における実証実験の、受容性評価の実施

n 国内・国際会議における
特別セッション

n 一般向けイベント実施

n メディア掲載:

新聞報道、テレビ報道多数

連携による成果

n 施策間での連携性の検討・実施

○ 実証実験データの相互利活用に向けた検討の実施 (情報大航海、センシングWeb)

n 他連携群 (次世代ロボット連携群) との連携の検討・実施

○ センシングWebでは、次世代ロボット連携群のインタフェース仕様を基に
サービス仕様記述、要求仕様記述を作成

(2) 法制度面の総括

制度・環境の整備

著作権

- n 著作権法改正に向けた提言やパブリックコメントなどによる継続的な働きかけにより、平成21年6月に、「**著作権法の一部を改正する法律**」が成立
 - その結果、検索サービス事業が可能となった。
- n **改正法の解釈の整理**
 - 事業者における改正著作権法の解釈を支援し、次世代サービスの展開の検討の手助けとする。
- n 情報産業における今後の次世代サービスの発展（産業活性化）と文化発展を実現する上で必要となる、**制度的問題 / 課題の抽出・検討**

個人情報保護

- n 扱いが不明確であった購買履歴や行動履歴などの個人に関連する情報も含めた**パーソナル情報の利活用のあり方の整理**を行っている。
- n パーソナル情報に係るシンポジウムの開催など、産学に対して**パーソナル情報の利活用に係る関心を喚起**した。
 - 産学主導によりパーソナル情報のサービス適用を検討するコンソーシアムの設立

(3) 情報発信の総括(1/3): ホームページ

n 情報発信の一環として連携群のホームページを立ち上げ、
国内外への積極的な情報発信を実施



(1) 日本語ホームページ



(2) 英語ホームページ

http://www.renkei.jst.go.jp/infor/infor.html

(3) 情報発信の総括(2/3): 国内講演、国際会議

- n 主監(西尾章治郎)が**国際会議等**で本連携施策群を紹介する**基調講演**を行う。
 - Second International Symposium on Universal Communication, December 2008.
 - International Conference on Complex, Intelligent and Software Intensive Systems, March 2009.
- n 『情報化月間2009』の**特別行事**において講演を実施
 - 定員を超える聴講者が来場するなど、高い関心を獲得
- n EUの主要検索エンジン関連プロジェクトであるChorusやOECDなどの開催する**国際会議**において講演を実施
 - EU各国の検索エンジン開発プロジェクトからの**技術連携要請**
 - 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ある**”との意見を頂くなど、国際的に高い評価を獲得



情報化月間2009 特別行事

2009年10月1日(木)
ANAインターコンチネンタルホテル東京



Chorus Final Conference

2009年5月26日(火)～27日(水)
ブリュッセル(ベルギー)



OECD Experts Conference

2009年6月8日(月)～9日(水)
リスボン(ポルトガル)

(3) 情報発信の総括(3/3): 国際会議、国内学会活動

人工知能学会全国大会 (2009年6月)、情報処理学会全国大会 (2010年3月予定) において**特別セッション**を設けるなど、幅広い分野への情報発信活動を行った。

防犯カメラ等に関する国際会議であるSPC2009 (International Conference on Security Camera Network, Privacy Protection and Community Safety, 2009年: 群馬で開催) において**特別セッション**を設け、センシングWeb技術の先導性を示した。

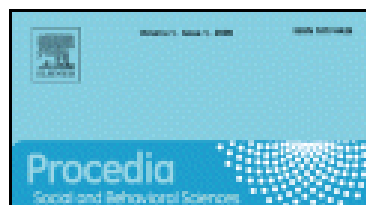
実証実験会場の商業施設において、**一般向けのイベント**を複数回実施した。

また、**新聞・テレビ**にも数多く取り上げられた。

(**新聞**: 朝日・産経・毎日・読売・日経・日刊工業等、**テレビ**: NHK・テレビ東京・KBS京都等)



人工知能学会全国大会



SPC2009



イベントの開催、
メディアへの情報発信

5 . 今後の展開

今後の連携施策群の進め方

情報爆発時代に新たな分野を切り拓くためには情報の利活用が必要

- Web上及び非Web上にある大量かつ多様な情報を、個人が簡便、的確、かつ安心して収集、分析できる次世代の知的情報アクセス基盤のさらなる推進
- 情報利活用する際の諸問題に対する調査、検討を行い、法制度面等も加味し、成果の効率的な活用が可能となるように活動を推進

今年度で終了する施策と継続する施策との連携体制が必要

- 今年度で終了する施策（情報大航海、センシングWeb）と継続する施策（情報信憑性検証技術、非順序型DB、Web社会分析）の議論の場を設ける。
- 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。