

「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

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2月4日

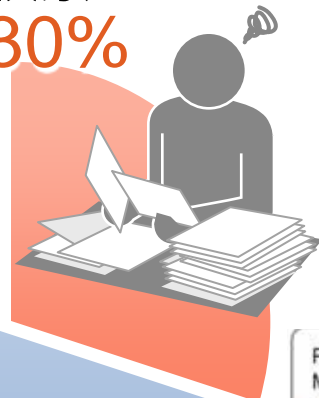
報告内容

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. 今年度の各省施策及び補完的課題における成果
4. 各省施策及び補完的課題の連携について
5. 今後の進め方

1. 背景・目的

情報爆発時代の到来：「探す」ことに明け暮れる

検索に費やす時間
30%



知的活動時間
(高付加価値産業従事者)

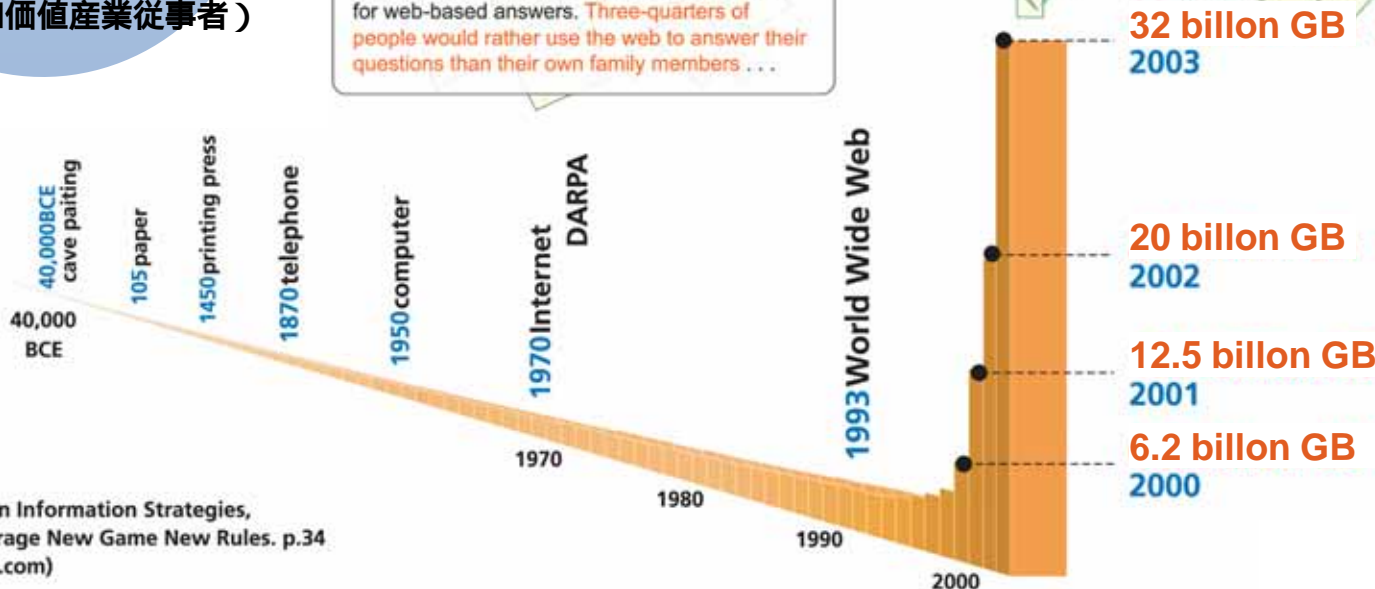
Non-1sec responseを目指す



複雑な質問の50%は答えが見つからない
75%の人々が家族よりもWebに尋ねる

情報爆発

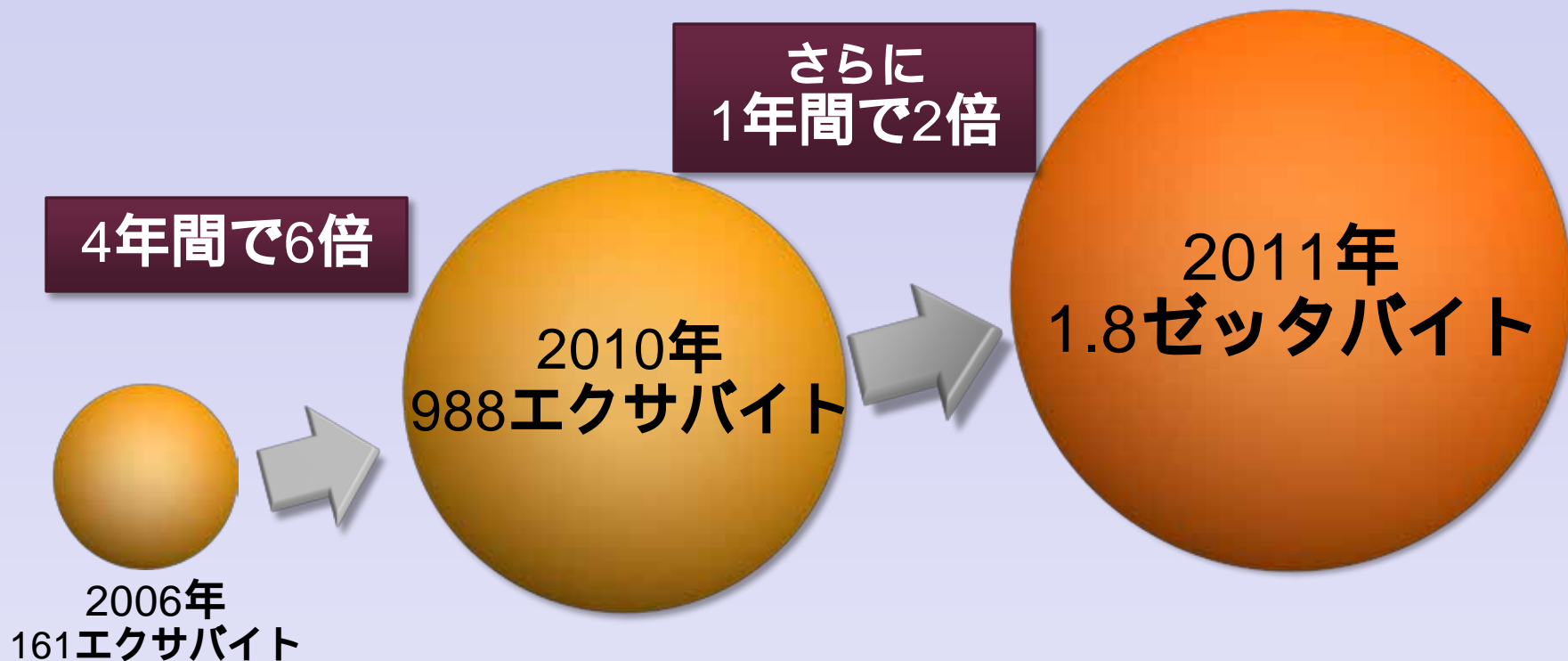
Plenty of information
MSN's own research found 50% of complex queries go unanswered, while even more simple searches take an average of 11 minutes. Despite this it found that there was real hunger for web-based answers. Three-quarters of people would rather use the web to answer their questions than their own family members...



Source: Horison Information Strategies, cited from Storage New Game New Rules. p.34 (www.horison.com)

今後の情報爆発時代

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

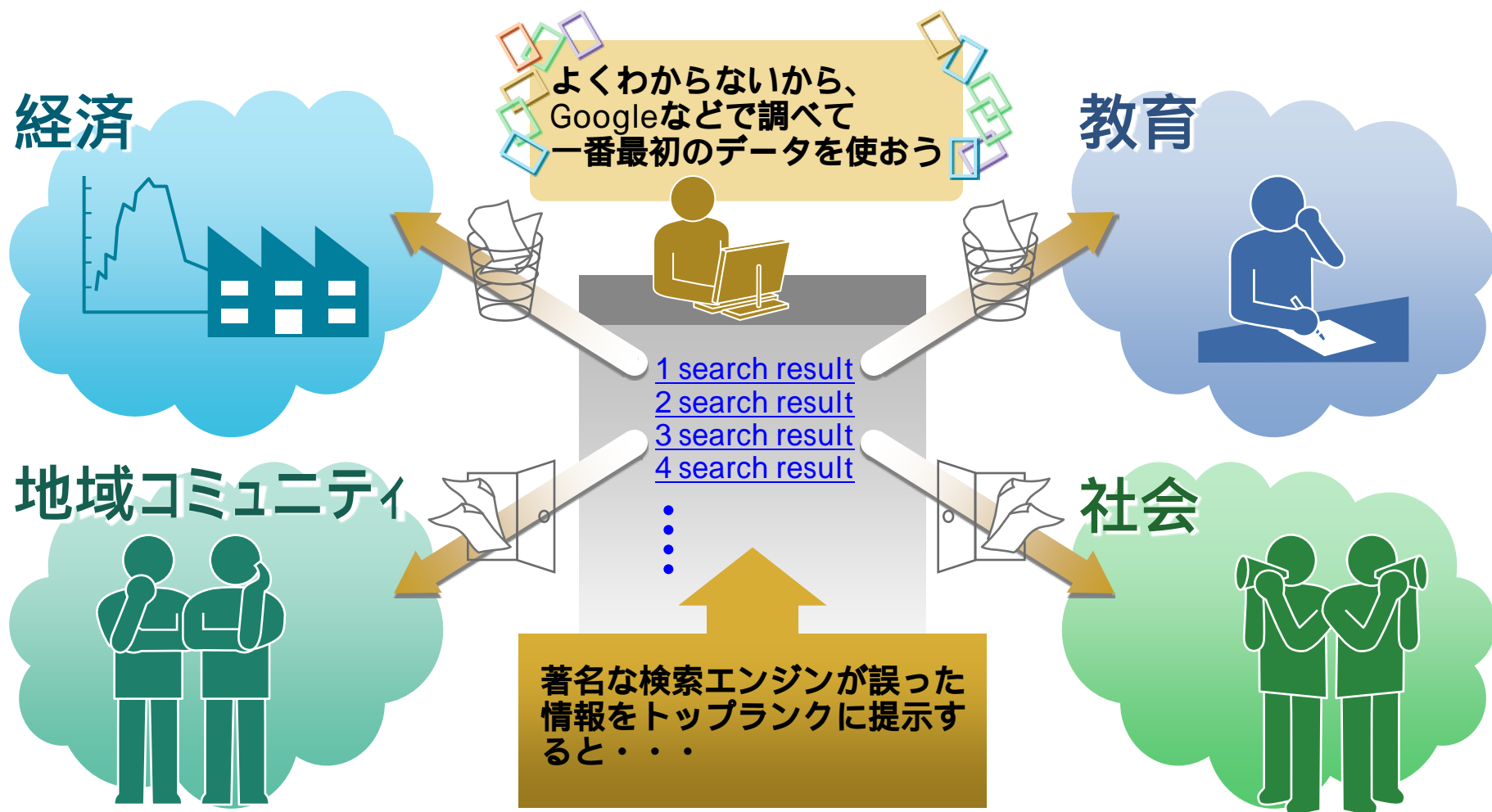


1エクサ(Exa)バイト = 10^6 兆バイト

1ゼッタ(zetta)バイト = 10^9 兆バイト

従来の検索サービスの社会的な問題の例

- 現在の検索エンジンは、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

玉石混淆のデジタル情報

経済



教育



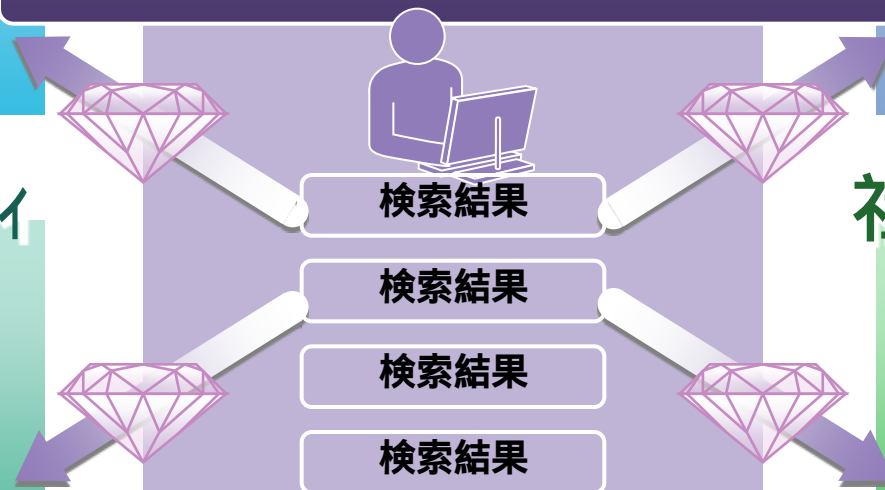
地域コミュニティ



社会

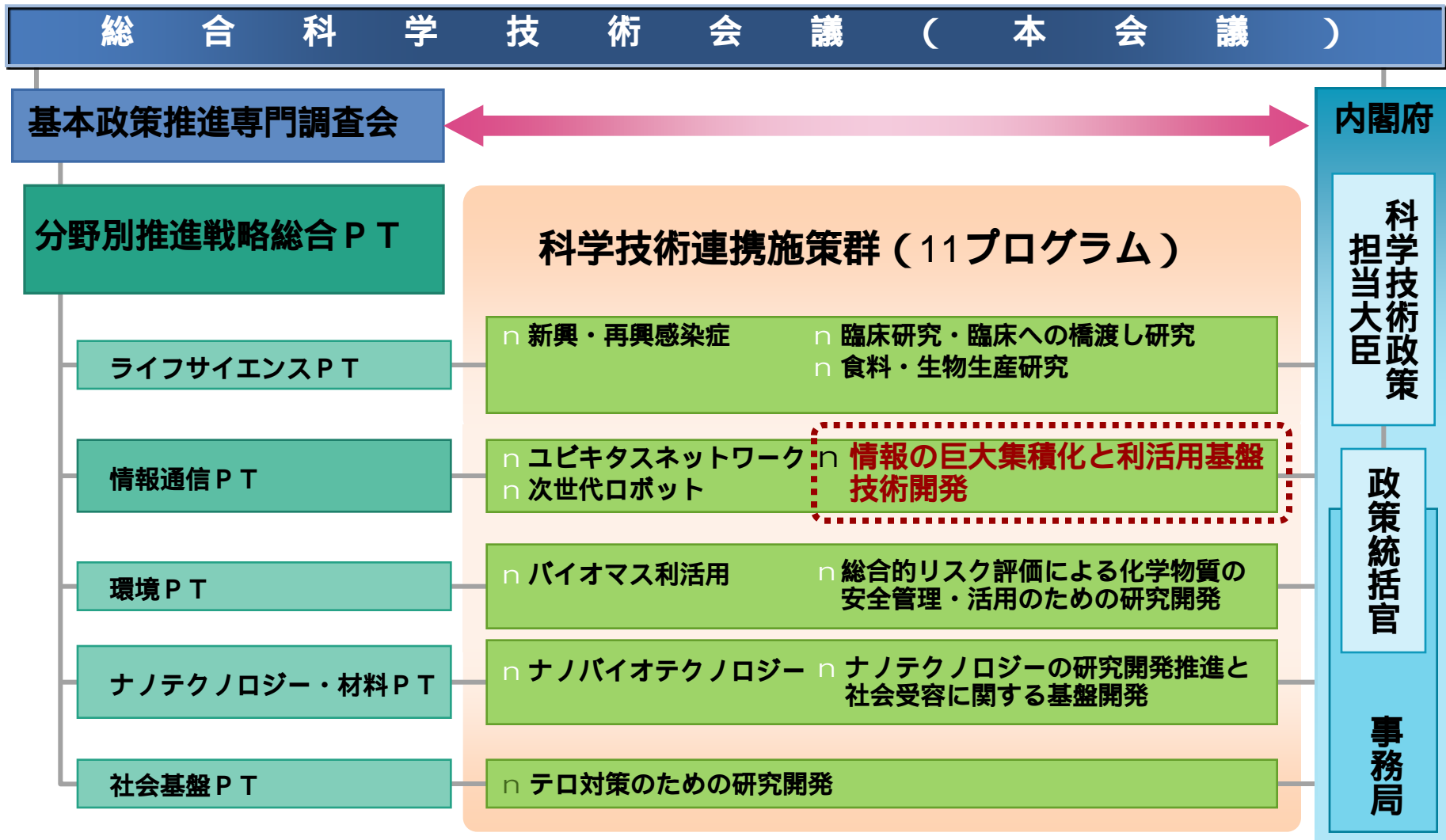


知的情報アクセス技術基盤
Ver.2011



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

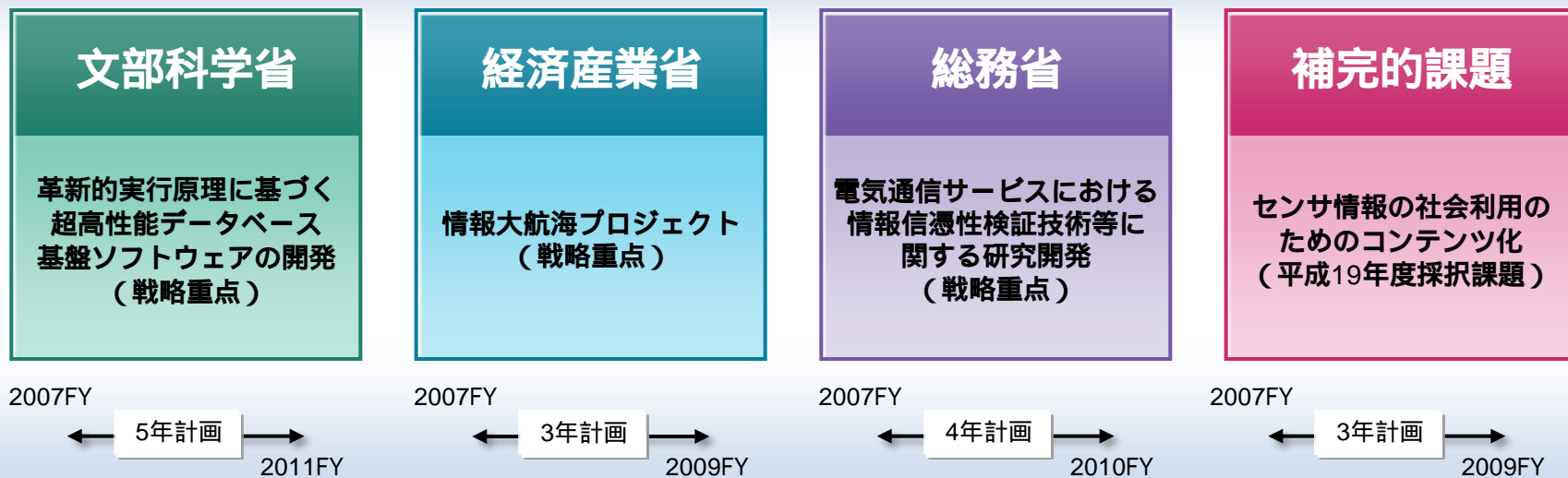


「情報の巨大集積化と利活用基盤技術開発」連携群

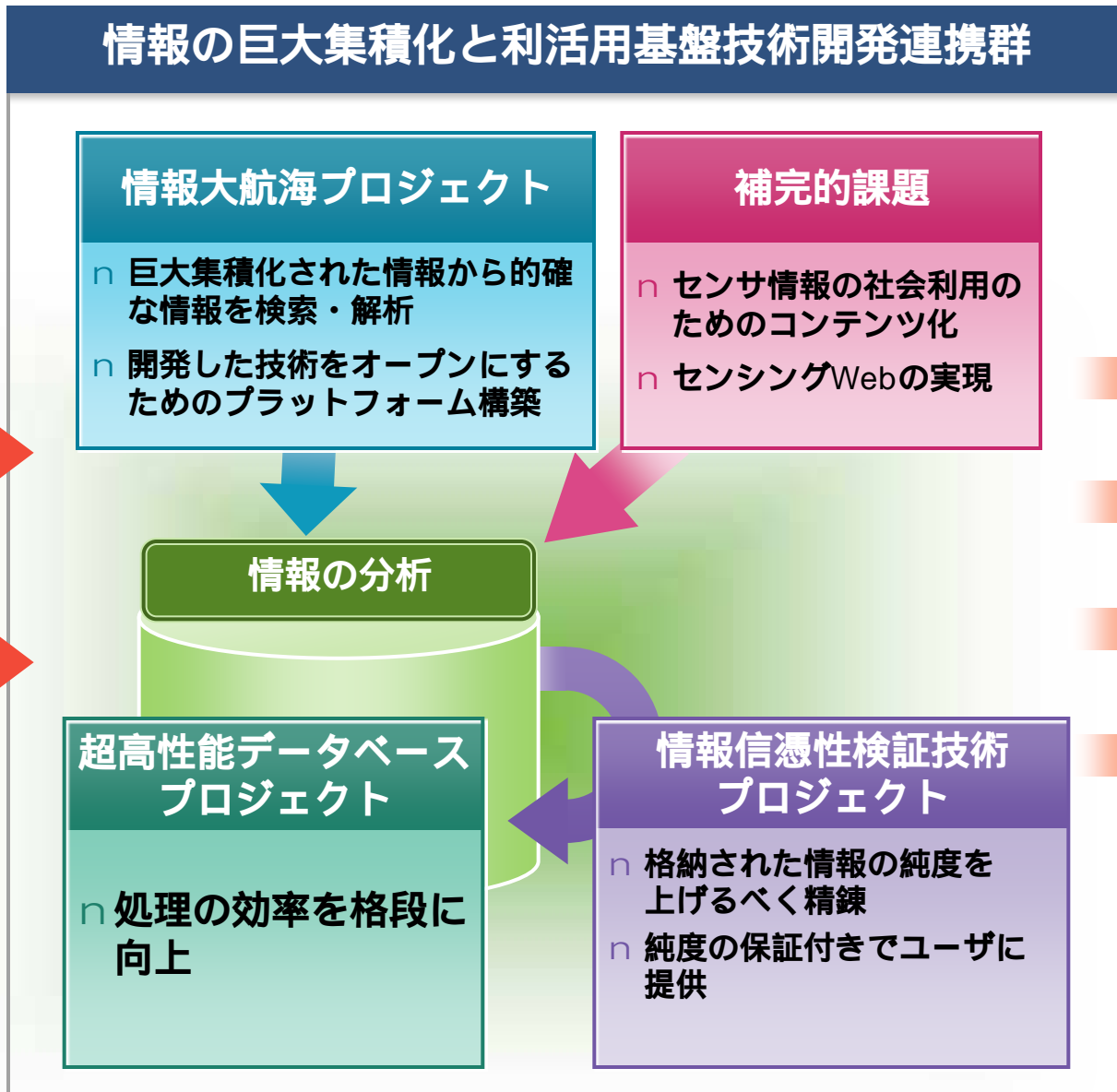
目標

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



各プロジェクトの位置づけ



2011年以降の「知的情報アクセス技術基盤」

P1: 情報大航海

次世代型 情報解析サービス

- 情報家電情報提供サービスAPI
- カーナビ情報提供サービスAPI
- 音楽コンテンツサービスAPI
- 検索エンジンAPI

従来型検索サービス

- xxx 検索エンジンAPI
- Web検索エンジンAPI

P2: 情報信憑性検証技術

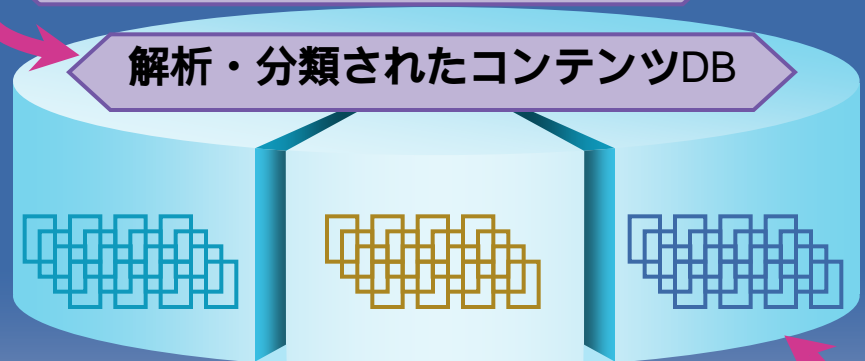
信憑性・信頼性分析サービス

P4: センシング Web

携帯検索エンジンAPI

信憑性・信頼性分析 API

解析・分類されたコンテンツDB

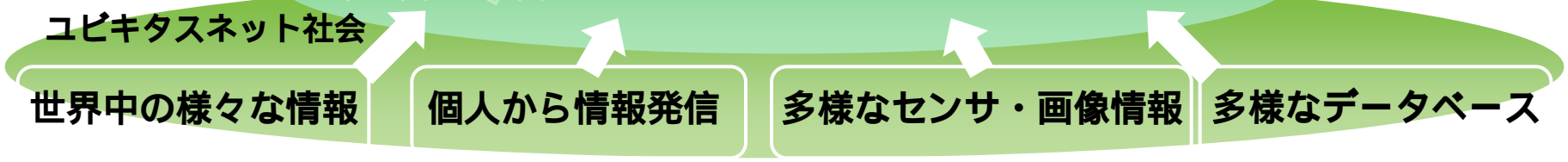


コンテンツの解析・分類・検索

P3: 超高性能データベース

情報融合炉の実現

インターネット上の多様なコンテンツ



3. 今年度の各省施策及び補完的課題 における成果

情報大航海プロジェクト

経済産業省

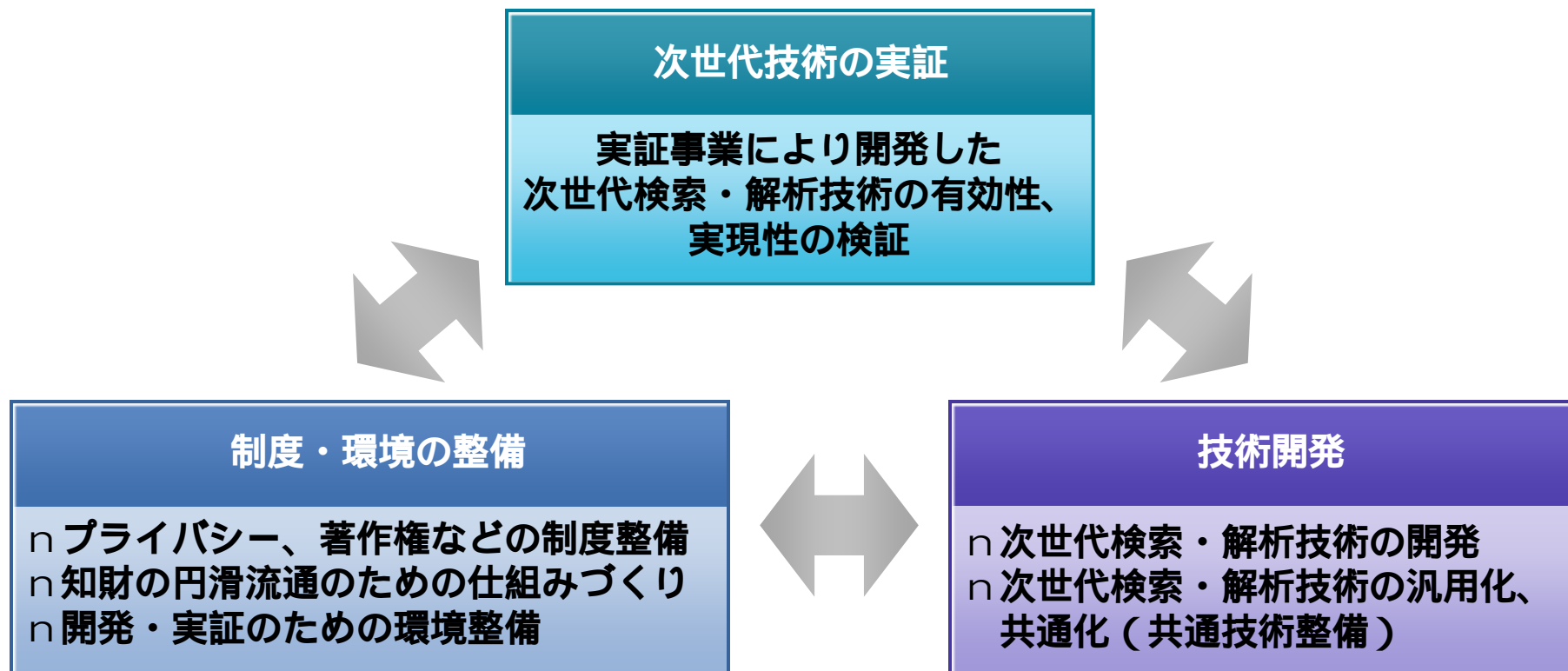
<http://www.igvpj.jp/index/>

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、より先進的な公的活動を含む事業へ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：次世代実証事業

次世代技術の実証

以下に示す3つの方向性から10の実証事業を実施し、イノベーション創出に向けての知見を得た。

A プライバシーに配慮した未来型サービス

ユーザのプロファイルや行動履歴などの情報をプライバシーに配慮して安全に収集・蓄積し、さまざまなサービスやコンテンツをマッチングして、ユーザ個人の生活スタイルにあわせた商品・サービス・情報の提供を行う。

個人情報（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

画像や映像の検索、対話型検索、感性検索などにより、従来のテキストベースでは得られなかったコンテンツへの直感的なアクセスや未知のコンテンツへのアクセスなどを開発することにより、コンテンツの新しい価値を見つけられる次世代のWebサービスを創出する。

情報検索等において、ユーザ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

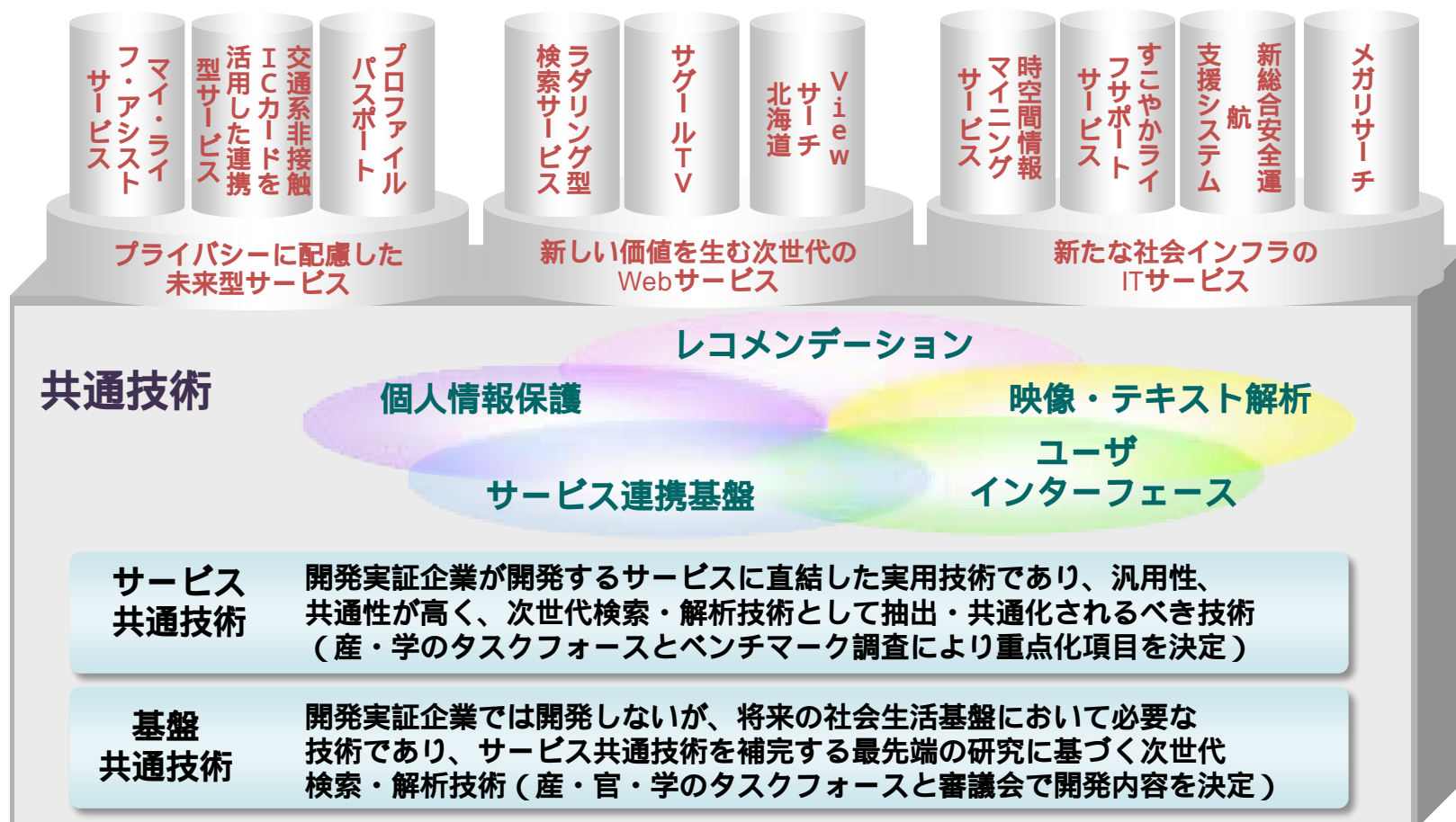
リアルタイム情報を含めた各種情報の統合的な解析により、各種の社会インフラにおける企業のリスク対策、医療情報の活用による高度医療サービスの提供など、ITを高度に活用し、より安心・安全な社会を実現する。

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。

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**を開発した。



今後の予定

次世代技術の実証

n 実証事業では、コラボレーション・プラットフォームを活用した次世代サービスの可能性の検証とコラボレーションの起点となる新たな共通技術等の創出を推進

共通技術

n 各分野における共通技術発展ロードマップを策定。国際標準化アクションプランを策定。共通技術アーキテクチャの改訂版を策定
n コラボレーションプラットフォームに関しては、利活用のための運用規定を見直し、より活用がしやすい方策を策定。CPの改良等による円滑な利活用環境の整備

制度面

n 個人情報匿名化技術等、開発している技術と制度方策の提案の連携を図る。これらにより個人情報ガイドラインやレコメンデーションガイドラインを策定
n 知財処理に関するケーススタディを形成し、知財処理阻組織の整備計画を作成。また知財バンクの運用に向けたアクションプランを作成

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

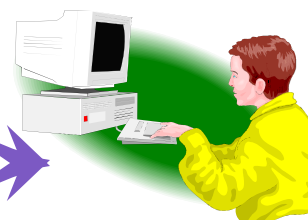
ネットワーク上の
玉石混淆な情報

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

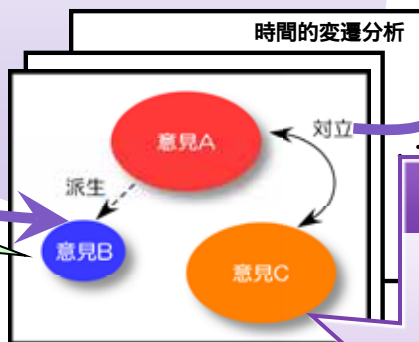
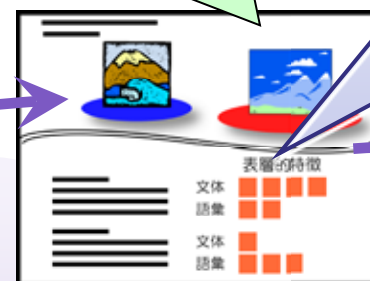


情報利用者がWebコンテンツの信憑性を判断する上での手がかりとなる情報を提供

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」

Webコンテンツの分析技術

1) Web上にある**マルチメディア情報の信憑性検証技術**

2) Web**テキスト情報の信憑性検証技術**

上記を開発し、周辺コンテンツや他情報源から収集したテキスト情報の分析・画像特徴量分析による信憑性検証等の基本的な技術を検証した。

1) Web上にあるマルチメディア情報の信憑性検証技術



Web**画像**の**典型度・過不足性・整合性**の分析

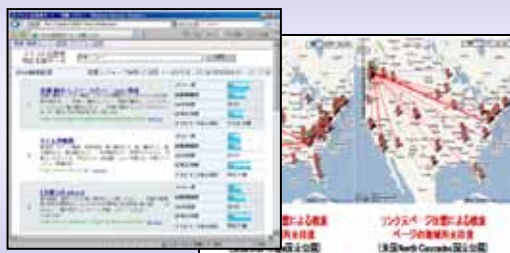


映像・音声の**偏り分析**や
視聴者**評価情報**の分析



Web**アーカイブ**を用いた**電子地図**
の**誤記載度・未記載度**の分析

2) Webテキスト情報の信憑性検証技術



検索ページの**内容のメジャー度**、
話題網羅度、**社会的な支持情報**の分析

発信者の**センチメント分析**、
熟知度分析

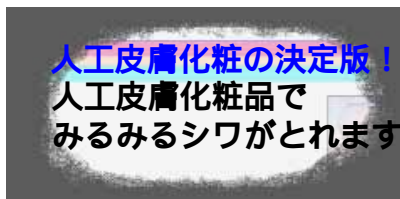


主な成果：「意味内容の時系列分析技術」

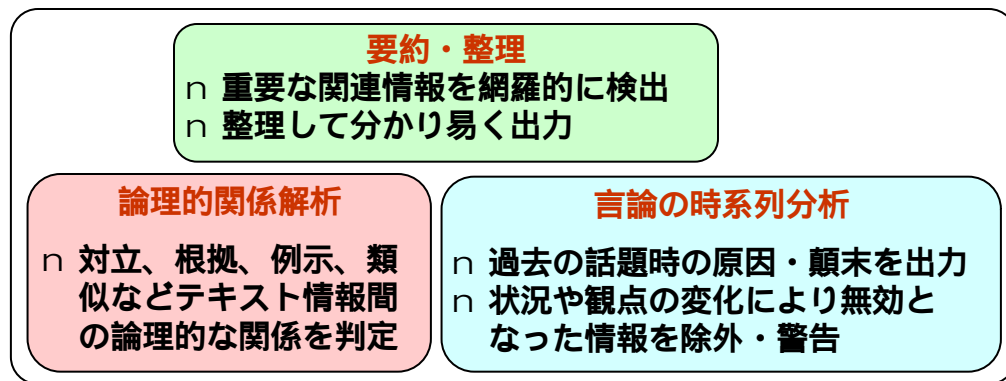
意味内容の時系列分析技術

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網羅的に提示し、**利用者の概観的な信頼性判断を支援する統合システム**構築に向けて、基本コンポーネントの研究開発に着手した。

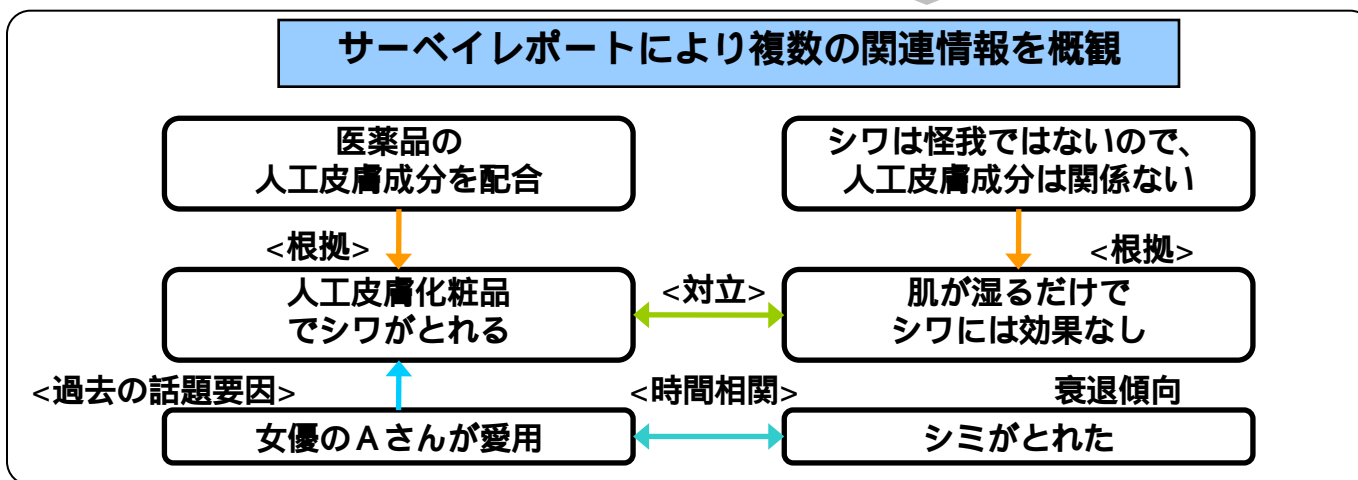
都合の良い情報だけを見ている
視野狭窄した状態



着目テキスト情報



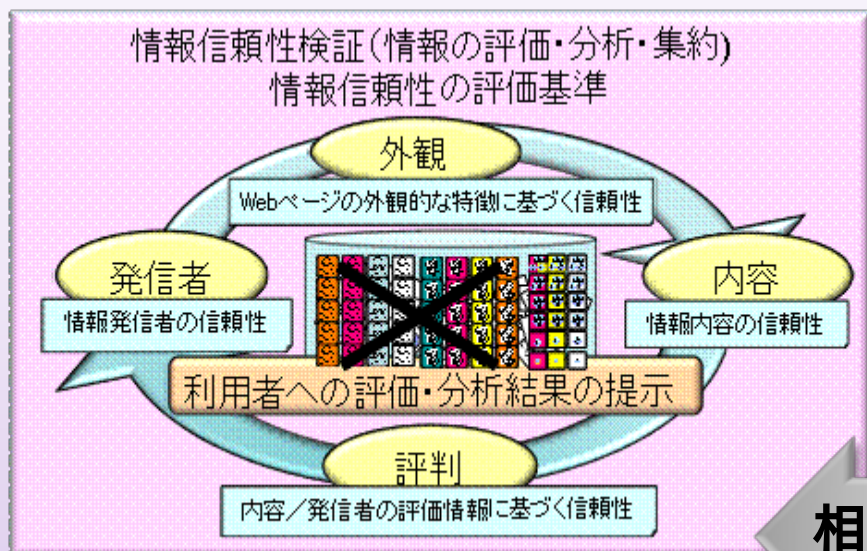
サーベイレポート



今後の予定：効果的な連携体制の構築

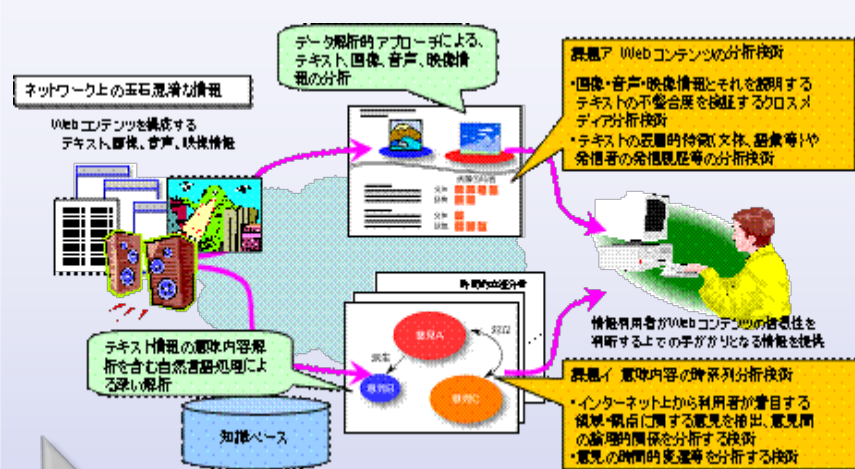
情報通信研究機構（NICT）自ら実施する研究と、委託研究の連携による体制を確立し、さらに情報大航海の基盤技術の利活用等も視野に入れ、誰もが安全・安心にWebコンテンツを利活用できるための情報信憑性検証技術を強力に推進する。

NICT研究



- n 大量のWebコンテンツから自然言語処理技術に基づく4種類の情報分析手法で、ユーザが信頼性を判断する基準となる情報を提供

委託研究



- n データ解析的アプローチによるWebコンテンツを構成するテキスト、画像、音声、映像情報の分析技術
- n 自然言語処理による深い解析による情報相互の論理的関係性、時間的変遷の分析技術

相互補完

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ークスルーが必要

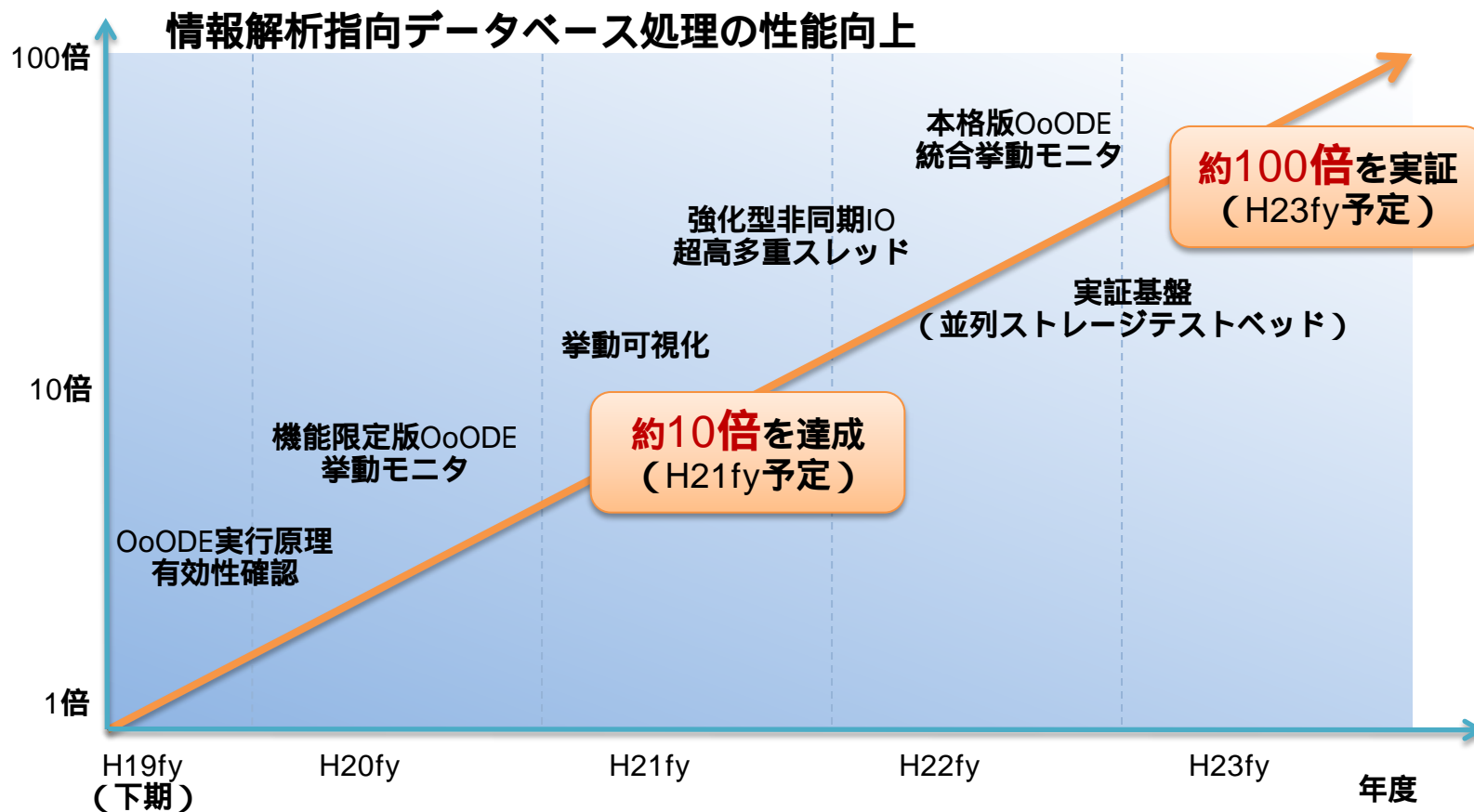
研究開発内容

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

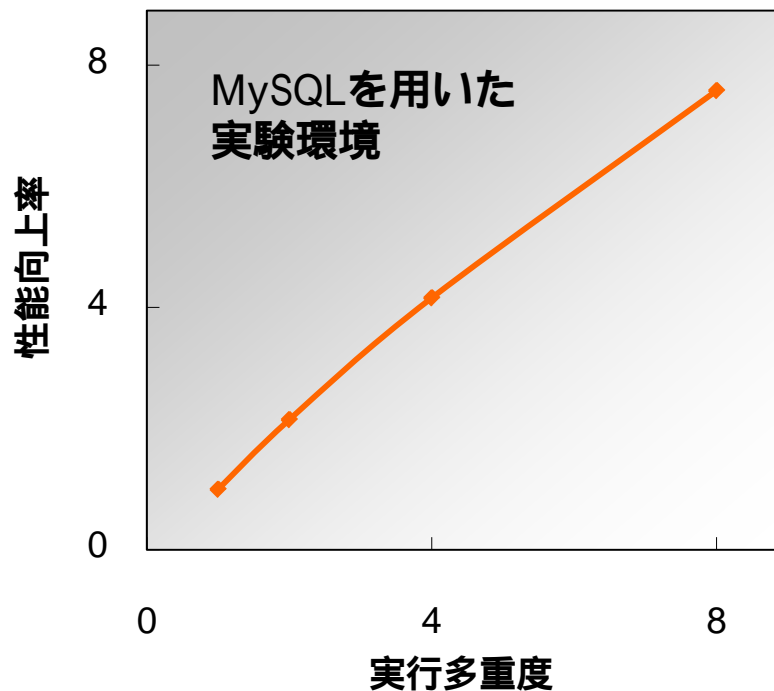
成果目標

- 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。

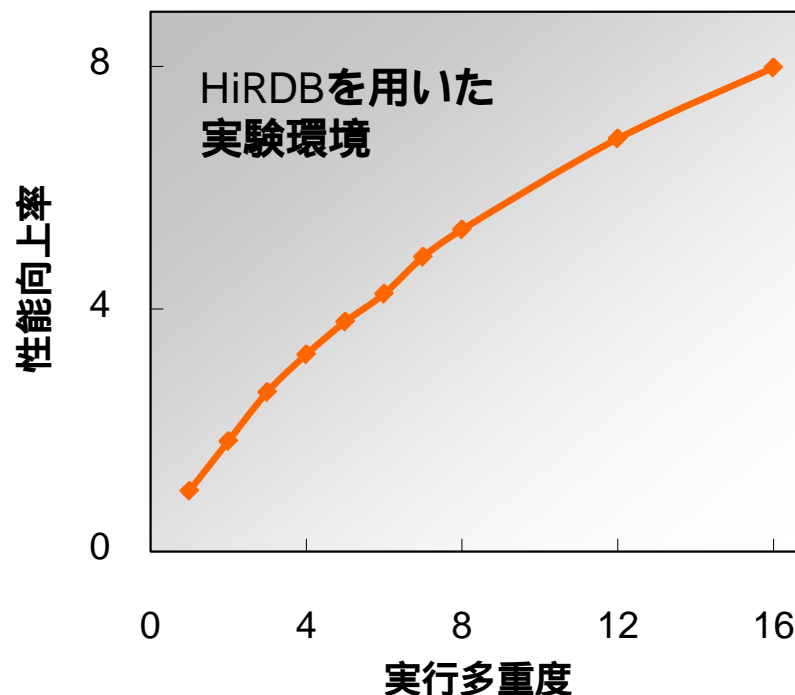


主な成果：実験結果

- n 既存のデータベースシステムを用いて、小規模実験を引き続き実施し、平成21年度の間目標である約10倍の性能向上に向けて、着実に進捗
- オープンソースのデータベースシステム (左図) では、実行多重度8に対して約8倍の性能向上を確認
 - 商用のデータベースシステム (右図) では、実行多重度16に対して約8倍の性能向上を確認



TPC-Hデータセット (SF=4.0; 4.0GB) Q.3
Itanium2 1.67GHz * 8, 30 HDDs, RedHat Linux AS3

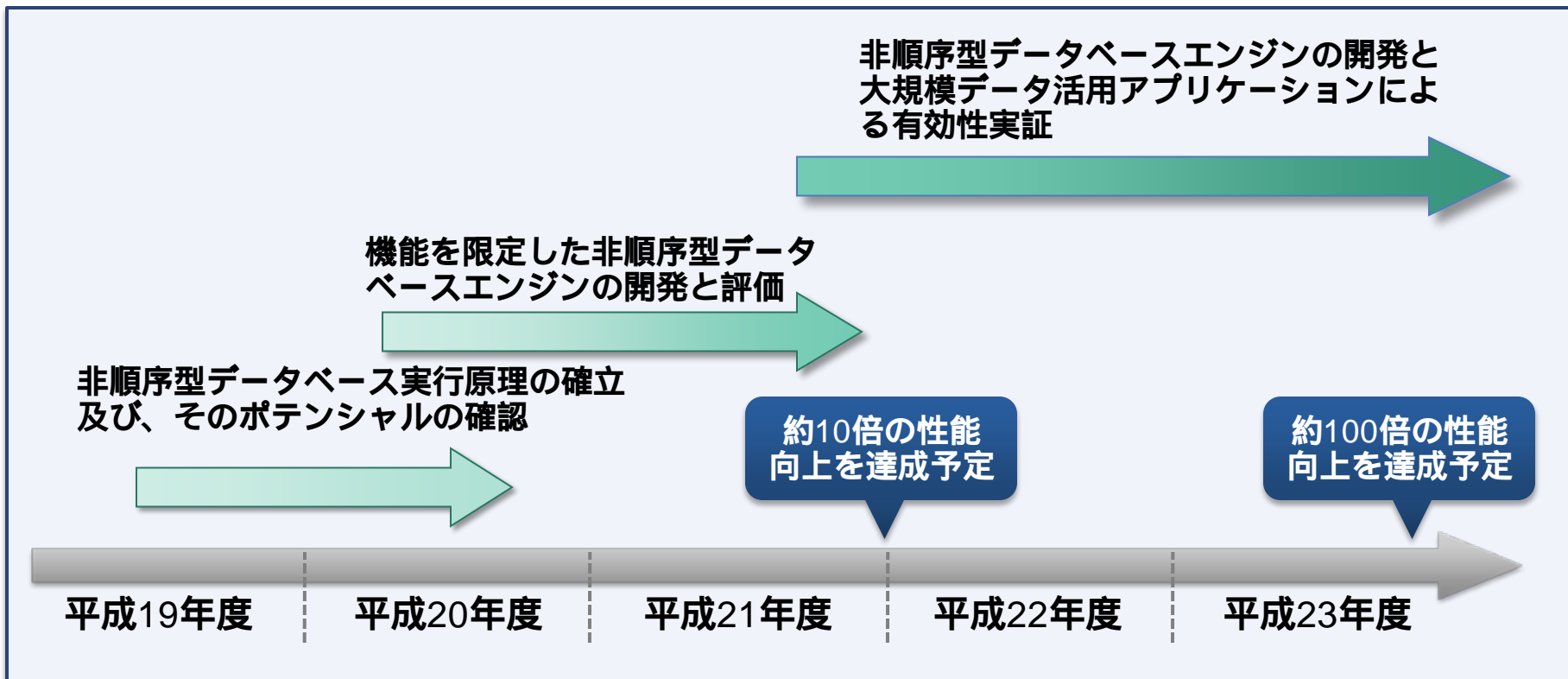


TPC-Hデータセット (SF=30; 30GB) Q.8
Xeon 2.40GHz, 1GB RAM, 10 HDDs, Fedora Core 4

今後の予定：性能向上に向けて

機能を限定した非順序型データベースエンジンの開発と評価

- ▶ 機能限定版の非順序型データベースエンジンの設計・実装
- ▶ 小規模実験により、非順序型データベースの性能向上を検証
- ▶ 超高多重非同期I/Oの処理を可能とする入出力制御機構の設計
- ▶ データベースエンジン、OS、ストレージの統合モニタリング機構の設計
- ▶ 実証評価基盤システムの設計及び一部を構築



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がたい
- 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、以下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)

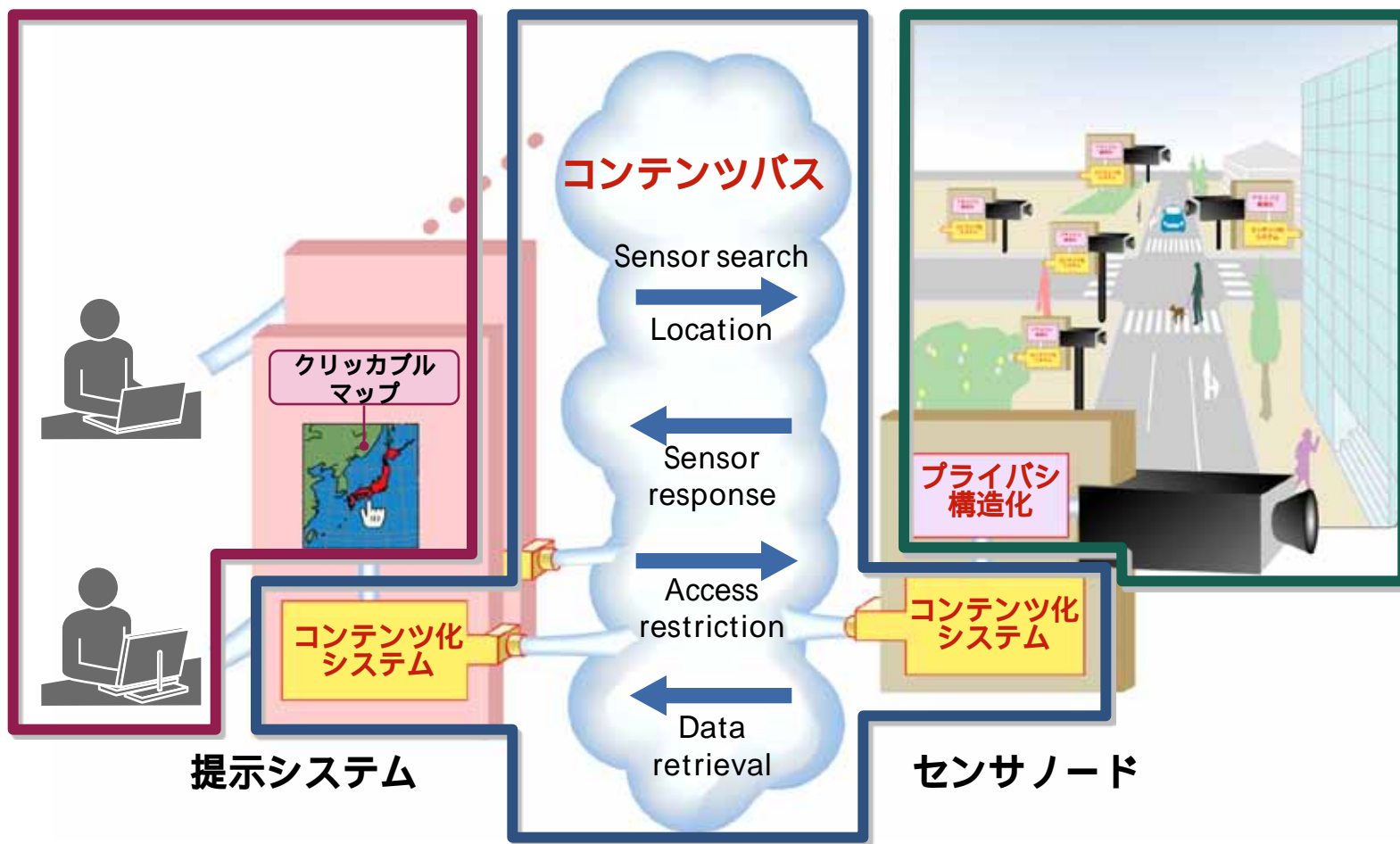
研究開発内容

センシングWeb

3 観測型実世界コンテンツ
提示技術（情報活用）

2 センサ情報コンテンツ化
技術（情報共有）

1 プライバシ情報管理技術
（アクセス管理）



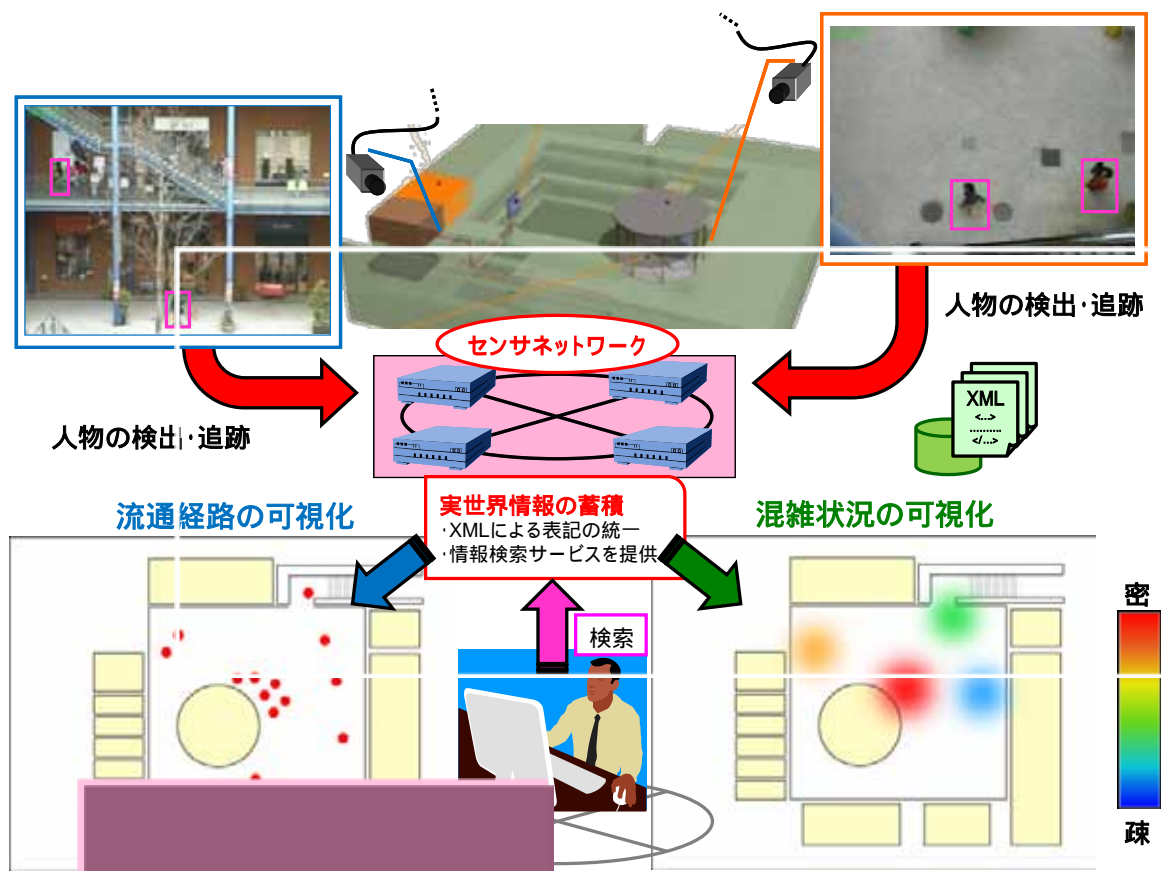
主な成果：センシングWeb技術開発

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



今後の予定：実証実験



各省施策及び補完的課題における総括： 主な成果一覧

主な成果一覧

	主な成果一覧	情報発信
<p>情報 大航海</p>	<p>n 次世代実証実験: 10サービスに関して実施</p> <p>n 共通技術: 映像・テキスト解析技術等、55の基本技術を開発 (http://www.igvpj.jp/index/cp-info/000/)</p> <p>n 制度面: ガイドライン作成 (個人情報保護、著作権等)</p>	<p>n 国内展示、 シンポジウム</p> <p>n 国際展示</p> <p>n メディア掲載: 34件</p>
<p>情報信憑性 検証技術</p>	<p>n Web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</p> <p>n Web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)</p>	<p>n 国内シンポジウム</p> <p>n 国際会議主催</p> <p>n メディア掲載</p>
<p>革新型DB</p>	<p>n 非順序型実行原理を小規模実験を実施し、約8倍の性能向上を確認 (OSS、市販DB等の実環境で検証)</p> <p>- 平成21年度の間目標は約10倍</p>	<p>n 国内シンポジウム</p>
<p>センシング Web</p>	<p>n センサ情報の記述法を開発 - プライバシーの構造化、バッファリング処理</p> <p>n 人物追跡データの公開方法</p>	<p>n 国内シンポジウム</p> <p>n 国際会議発表</p> <p>n 国内学会誌特集号掲載</p>

情報発信の例：国内展示、国際講演（情報大航海）

- n 幅広いユーザから大きな関心をもたれている最先端IT・エレクトロニクス総合展CEATEC JAPAN2008に、昨年引き続き出展し、**全般において非常に高い関心**を集め、メディアにも多く取り上げられ、大変盛況であった。
- n EUにおけるICT関連の主要イベントであるICT2008における招待講演（喜連川教授）を実施。予定していた以上に講演を依頼される等、日本の活動に対する関心は非常に高いものであった。

CEATEC JAPAN 2008への出展

- n 会期: 2008年9月30日(火) ~ 10月4日(土) 於: 幕張メッセ
- n 概要:
 - ・ブース展示 (NTTドコモ等、11団体のデモ展示)
 - ・コラボレーション・コーナー(実証事業等の説明)
 - ・メインステージ(全体概要の説明)
 - ・キーノートスピーチ(喜連川教授による講演: 500人聴講)
 - ・セミナー(制度・知財に関するパネルディスカッション: 120人聴講)
- n ブース来場者数: 延べ約17,000人(5日間合計)
オーストラリア大使館商務官をはじめ、多くの企業等から開発した共通技術の活用やコラボレーションについて問い合わせがなされる。
- n メディアへの掲載: 『日経BP』 『IT media』 『東洋経済』 『日刊工業新聞』 『電波新聞』 他



EU-IUT2008での講演

- n 会期: 2008年11月25日(火) ~ 11月27日(木) 於: リオン(フランス)
- n 登壇発表: CHORUSミーティングでの講演
 - “Prosip Disruptive innovation inでの講演 time ch”
 - “Networkæ&3D Inteでの講演”
- n コネクション: 欧州委員会のキーマン(Head of Unit)からの上記講演に対する謝辞
NESSI (Networked European Software and Services Initiative) との継続的な情報交換
QUAERO参加企業である、exalead社訪問



情報発信の例：国際会議、国内学会活動 (情報信憑性、センシングWeb)

- 国際ワークショップ (WICOW 2008: Second Workshop on Information Credibility on the Web, Napa Valley, California, October 30th, 2008) を企画開催し、情報信憑性技術を紹介するとともに、先導性を示した。
- パターン認識分野最大の国際会議であるICPR (International Conference on Pattern Recognition, 2008年: Florida, Tampaで開催) においてワークショップを設け、センシングWeb技術の先導性を示した。
- 人文科学や社会科学も含めた学際的な学会誌である人工知能学会誌において「センシングウェブ」特集号が掲載 (2009年3月号)



『特集「センシングウェブ」にあたって』：
馬場口登 (阪大), 美濃導彦 (京大)

『センシングウェブ～概念と課題～』：
美濃導彦 (京大)

『連携施策群におけるセンシングウェブ技術の重要性』：
西尾章治郎 (阪大)

『カメラ映像におけるプライバシー対策のためのアプローチ』：
美濃導彦 (京大) 他

『音声に含まれるプライバシー情報の保護』：
中川聖一 (豊橋技科大) 他

『センサ情報共有のためのアーキテクチャとデータ形式標準』：
鯉坂恒夫 (和歌山大) 他

『センシングウェブにおけるセンサ情報の構造化』：
谷口倫一郎 (九大) 他

『シースルービジョンとプライバシー情報の管理』：
大田友一 (筑波大) 他

『観測型実世界コンテンツ：デジタルジオラマ』：
馬場口登 (阪大) 他

<http://www.dl.kuis.kyoto-u.ac.jp/wicow2>

(1) WICOW2008の企画開催

(2) ICPR2008ワークショップ

(3) 人工知能学会特集号

4. 各省施策及び補完的課題の連携について

「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

電気通信サービスに関する情報信憑性・ 信頼性技術等に関する研究開発（総務省）

Webコンテンツの分析技術

- ・ 画像・音声・映像情報の分析技術
- ・ テキスト情報の分析技術

意味内容の時系列分析技術

情報発信者の信頼性評価技術

Webページの外観的特徴に基づく信頼性評価技術

情報内容に基づく信頼性評価技術

内容 / 発信者の評価情報に基づく信頼性評価技術

情報大航海プロジェクト（必要な情報を簡便かつ的確 に検索・解析するための技術開発）（経済産業省）

コラボレーションプラットフォーム

超高速パターン発見技術

レコメンデーション技術

- ・ 状況に柔軟に対応するレコメンデーション技術

- ・ 複数のレコメンデーションの統合技術

画像・映像の意味理解技術

広範な日本語表現解析技術

対話による検索結果絞込み

位置データ統合基盤技術

個人情報匿名化技術

画像分類結果可視化技術

センサーデバイスのゲートウェイ技術

連携

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）

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

- ・ 非順序型データベースエンジン技術に関する研究
- ・ 資源調整技術に関する研究
- ・ モニタリング技術に関する研究
- ・ 実証評価に関する研究

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）

プライバシー情報管理技術

センサ情報コンテンツ化技術

観測型実世界コンテンツ提示技術

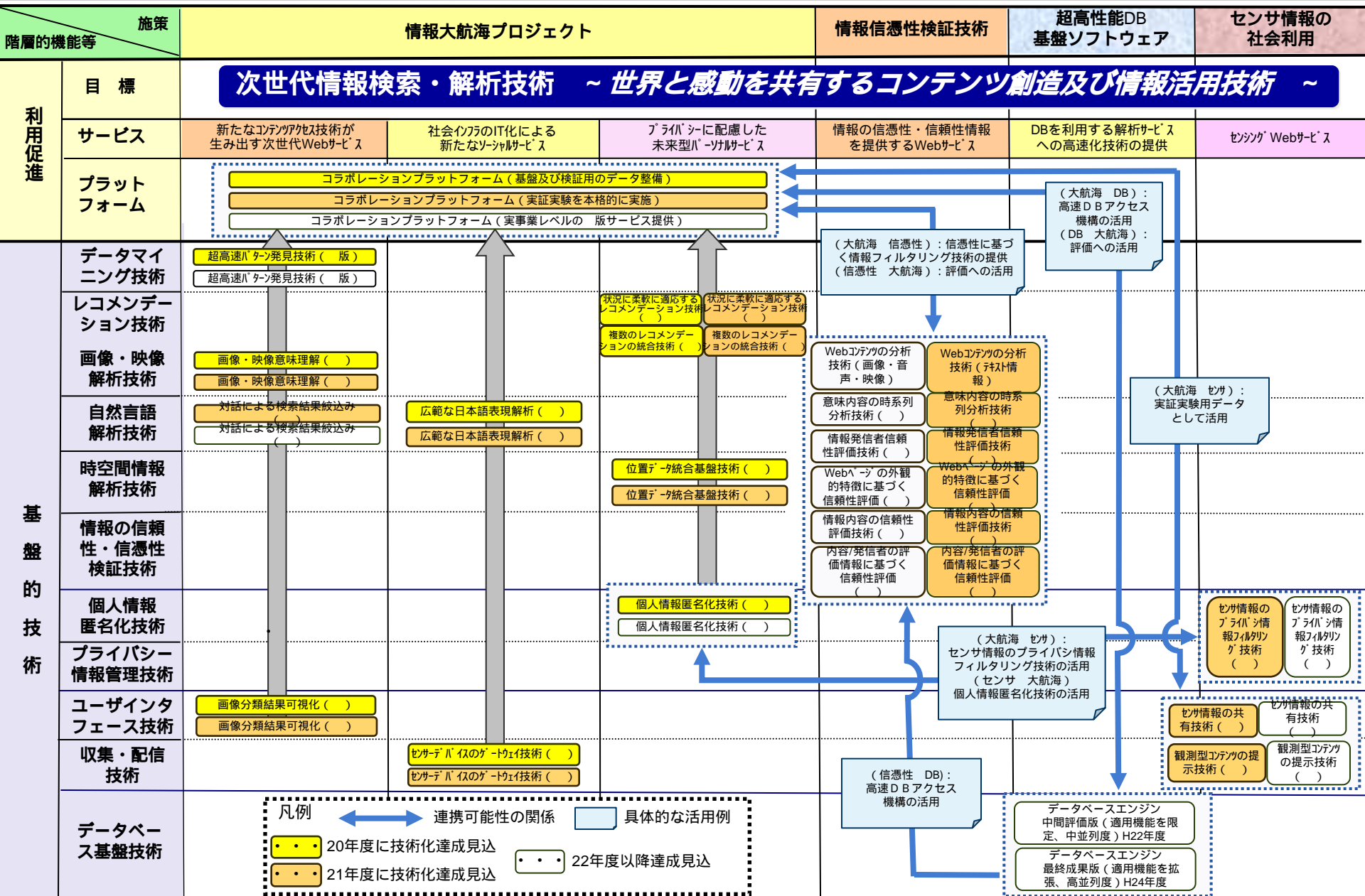


目標

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。

「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）



5. 今後の進め方

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

積極的に情報発信を推進するとともに一層の連携体制の強化が必要

- ▶ シンポジウム開催等による情報発信や専門家との議論の場を設ける。
- ▶ 今後の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。

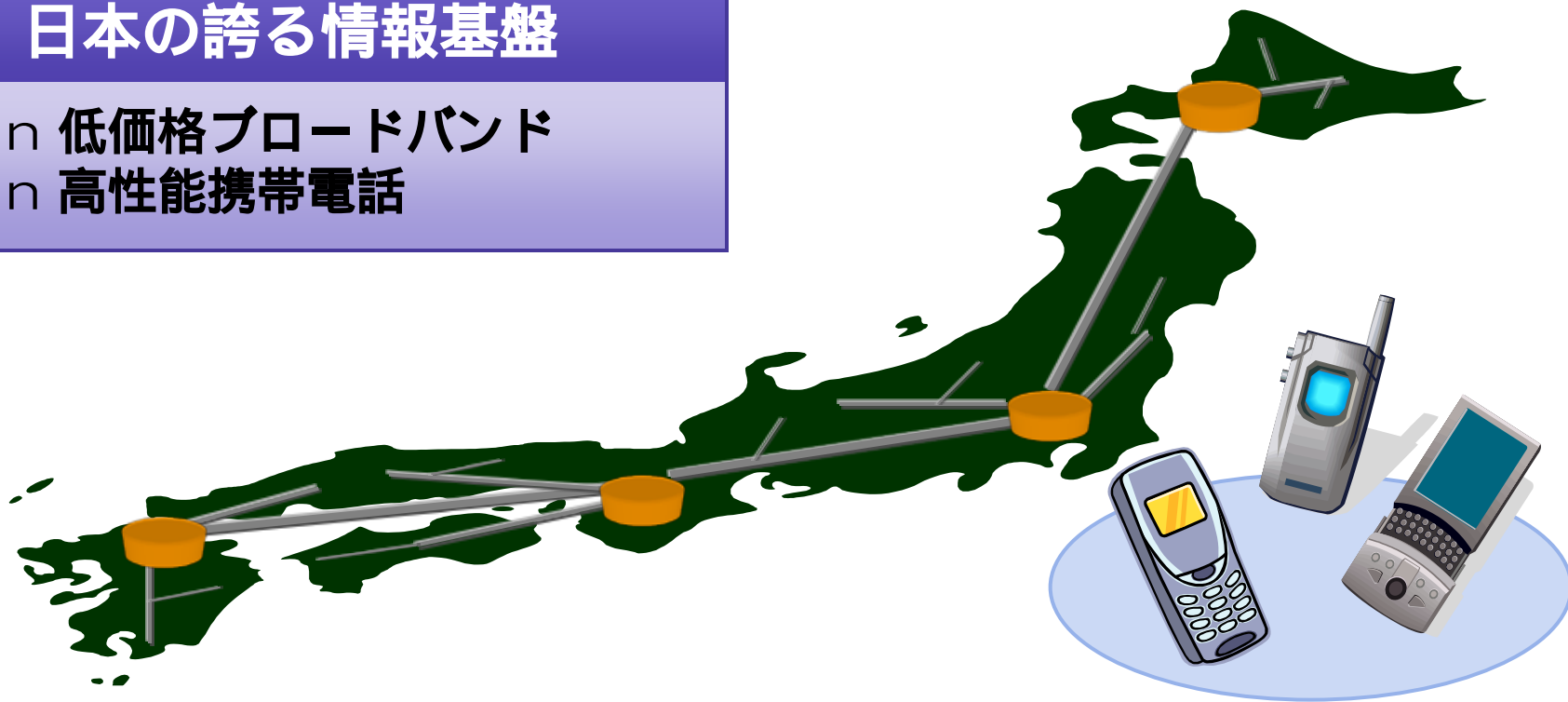
「知の泉」から「知を紡ぐ」



世界をリードする日本の情報基盤

日本の誇る情報基盤

- n 低価格ブロードバンド
- n 高性能携帯電話



我が国には、
「知を紡ぐ」プラットフォームの構築に向けて
リーダーシップを発揮する宿命がある！