

# 情報爆発(Info-plosion)プロジェクト

2010年10月1日 14:00- @国立情報学研究所

**代表 喜連川優**

**東京大学 教授**

**東京大学地球観測データ統合連携研究機構長**

**東京大学生産技術研究所戦略情報国際研究センター長**

**New IT Infrastructure for the Information Explosion Era**

Research Project funded by MEXT Grant-in-Aid for  
Scientific Research in Priority Areas (FY 2005-2010)

*info*-plosion

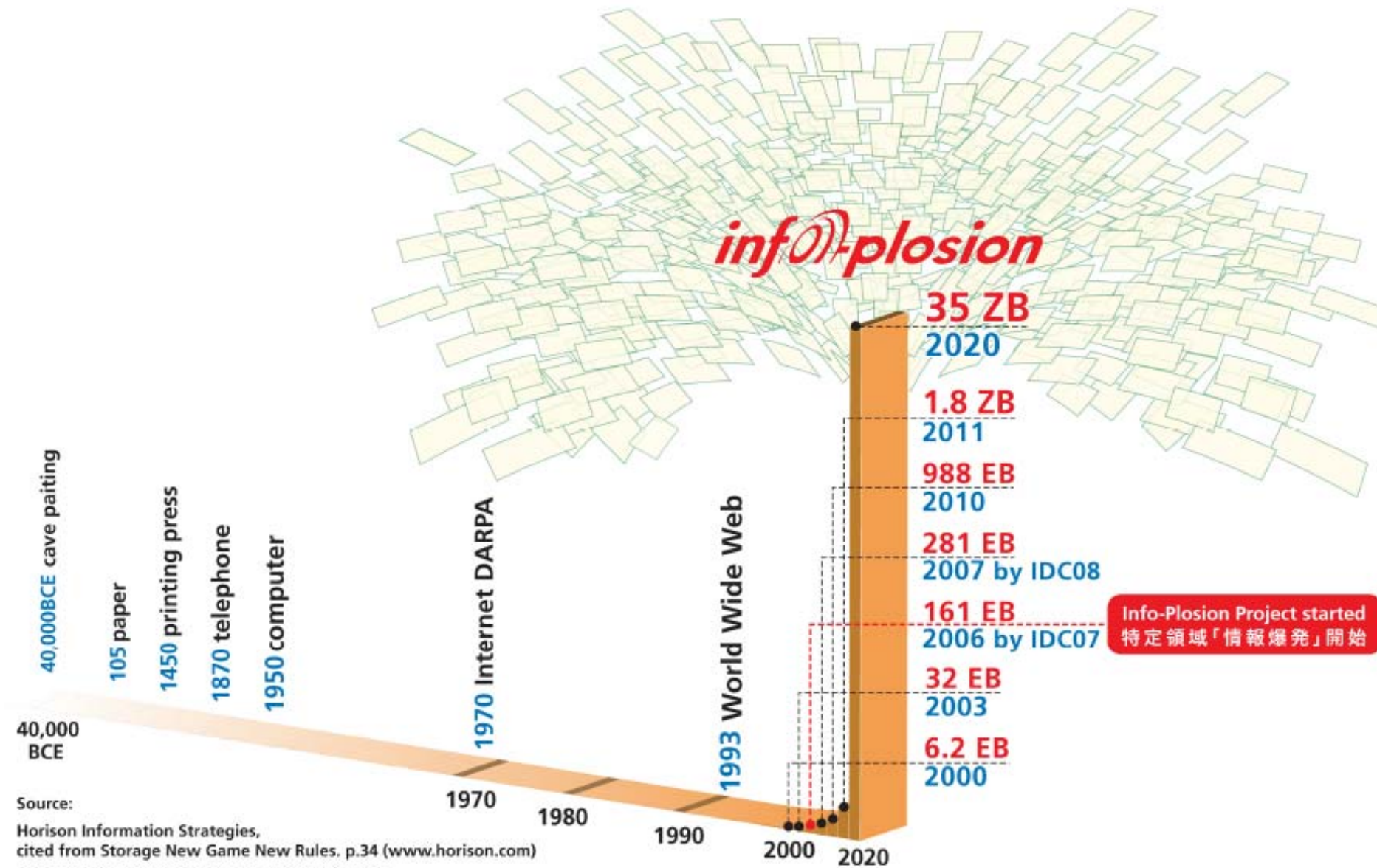
# 情報爆発時代に向けた新しいIT基盤技術の研究

平成17年発足 文部科学省特定領域研究

(2005.7 – 2011.3) 2004年申請

喜連川優

東京大学 生産技術研究所



Source:  
Horison Information Strategies,  
cited from Storage New Game New Rules. p.34 (www.horison.com)  
IDC: The Diverse and Exploding Digital Universe  
2020: <http://www.emc.com/collateral/demos/microsites/idc-digital-universe/iview.htm>

# 現代用語の基礎知識掲載

サーチエンジンを使って見る情報はほんの一握り

サーチエンジンが重要な情報と判断したページ

上位10件

多数派意見

恣意的なランク操作によって上位に浮上

自由国民社 2007.1

## 現代用語の基礎知識

Encyclopedia of Contemporary Words

100万人の生活スタイル

2007年の各分野いまがみえる100枚のわかる絵

最新カタカナ・略語辞典

マナーと常識事典

日本人なら感じておきたい情報爆発時代の心がまえ

日本人の不安にこたえる正規分布と格差社会

用語収集大作戦

伊藤園 お茶

Hatena

2007

since 1948

自由国民社

### 情報とサーチ

知的集約度の高い職業に従事する人々は仕事時間の3割を必要な情報を「探す(サーチ)」という行為に費やしているという報告がある。情報爆発時代においては、この探すという行為を的確に、かつ素早く実現する技術が大切になり、その開発が大きな挑戦でもある。

情報とは、

例えば、情報サーチエンジンがされている。それが普通なる。残りの見ることが普通でなく困難を順序づけ

### 情報の

情報を探す」という多くの時間を費やしている。

量が增大する、いわば、「情報爆発時代」を迎えるに技術の出現により、きわめて容易に全世界の人々に情報を発信することが可能となった。

しかしながら、ふと気がつく、情報は膨れ上がり、自分が真に必要とする情報を探すことが必ずしも簡単ではない。

情報を探す」という多くの時間を費やしている。

量が增大する、いわば、「情報爆発時代」を迎えるに技術の出現により、きわめて容易に全世界の人々に情報を発信することが可能となった。

しかしながら、ふと気がつく、情報は膨れ上がり、自分が真に必要とする情報を探すことが必ずしも簡単ではない。

## 日本人なら感じておきたい 情報爆発時代の心がまえ

本当に欲しい情報、信頼できる情報に到達するために

東京大学教授 喜連川 優

2003

東京大学教授、生産技術研究所 略情報融合国際センター長。1983年東京大学工学系研究科情報工学専攻博士課程修了。工学博士。

データベース工学、ウェブマイニング、ストレージシステム、デジタルアースなどの研究に従事。

### 情報爆発

検索にかかる時間

30%

実質的な活動時間

知的集約度の高い仕事に従事する人は、活動時間の30%は「欲しい情報を探す」ことに明け暮れている。

2002

2001

情報爆発的に増大しても、一人の人間が読める量は以前と変わらない。

1993 2000

World Wide Web

1870 電話

1950 コンピュータ

1970 インターネット

1450 活版印刷

40,000BCE 洞窟壁画

105 紙

情報量

### 特集 日本人なら感じておきたい 情報爆発時代の心がまえ

これからのデジタル世界と周辺キーワード

ロングテール

従来多くの事象はパレートの法則つまり20%の商品が売り上げの80%を生み出すという80

この情報は簡単に見つけ出すことのできる仕組みを考え出すことは人類にとっての新たな課題となっている。

イラスト：情報設計 (info-architects.jp)

# 計画研究班

## A01 情報爆発時代における情報管理・融合・活用基盤

「情報爆発時代におけるサイバー空間情報定量評価基盤の構築」喜連川優（東大）

「情報爆発時代の情報検索基盤技術」安達淳（NII）

「情報爆発に対応するコンテンツ融合と操作環境融合に関する研究」田中克己（京大）

「ユビキタスネットワークコンテンツに対する管理・統合基盤に関する研究」下條真司（NICT）

「構造的言語処理による情報検索基盤技術の構築」黒橋禎夫（京大）

## A02 情報爆発時代における安全・安心ITシステム基盤

「情報爆発に対する高度にスケーラブルな高性能自律構成実行基盤」松岡聡（東工大）

「情報爆発に対応する高度にスケーラブルでセキュアなソフトウェア構成・更新方式」柴山悦哉（東大）

「情報爆発に対応する高度にスケーラブルなソフトウェア構成基盤」近山隆（東大）

「情報爆発に対応する高度にスケーラブルなモニタリングアーキテクチャ」中島達夫（早稲田大）

## A03 情報爆発時代におけるヒューマンコミュニケーション基盤

「相互適応可能な実世界インタラクションのための計算モデル・システムの構築」松山隆司（京大）

「実世界インタラクションの分析・支援・コンテンツ化」西田豊明（京大）

「セマンティック・ヒューマノイドインタラクション」國吉康夫（東大）

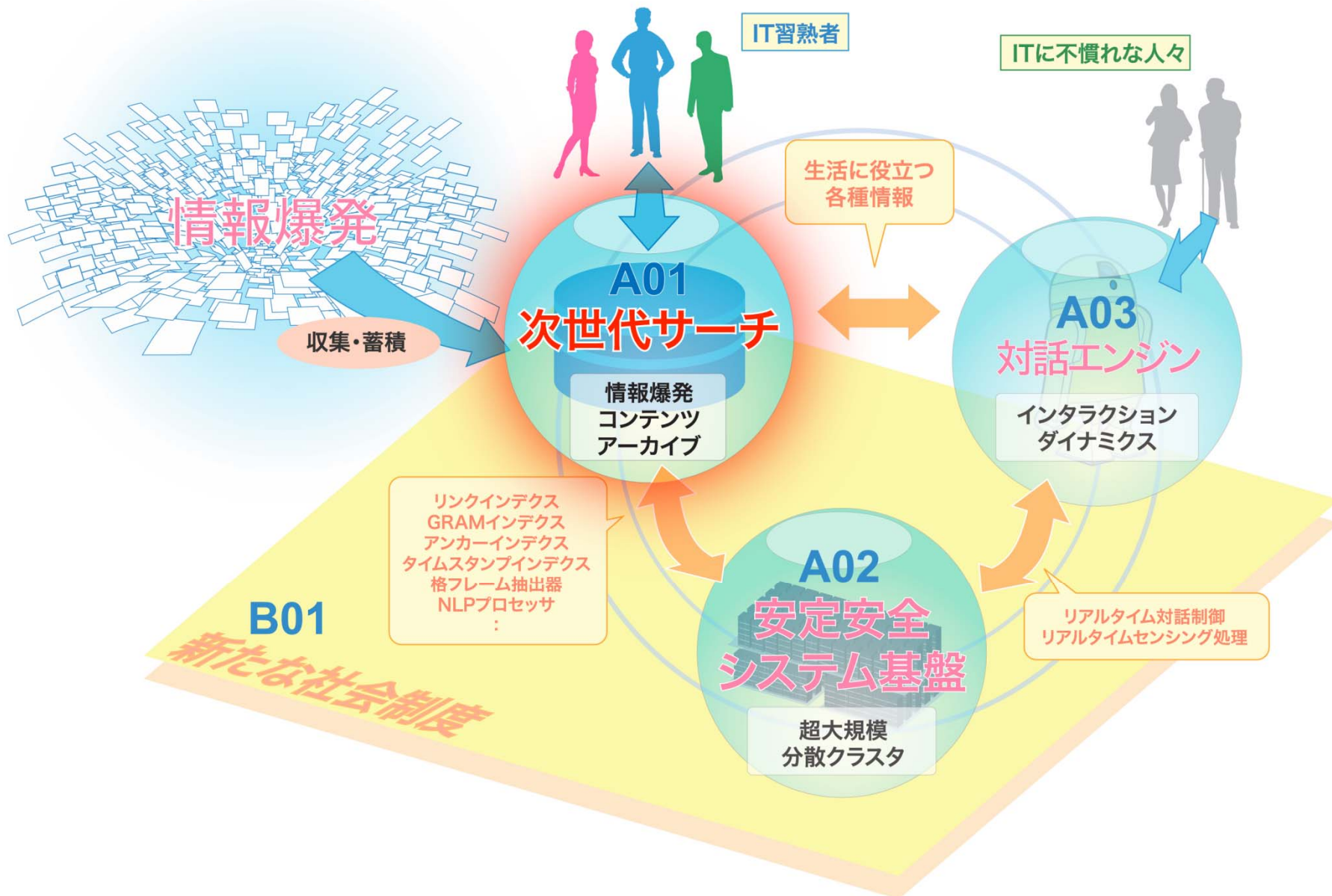
## B01 情報爆発時代における知識社会形成ガバナンス

「知識社会経済システムの競争的発展とそのガバナンスに関する研究」須藤修（東大）

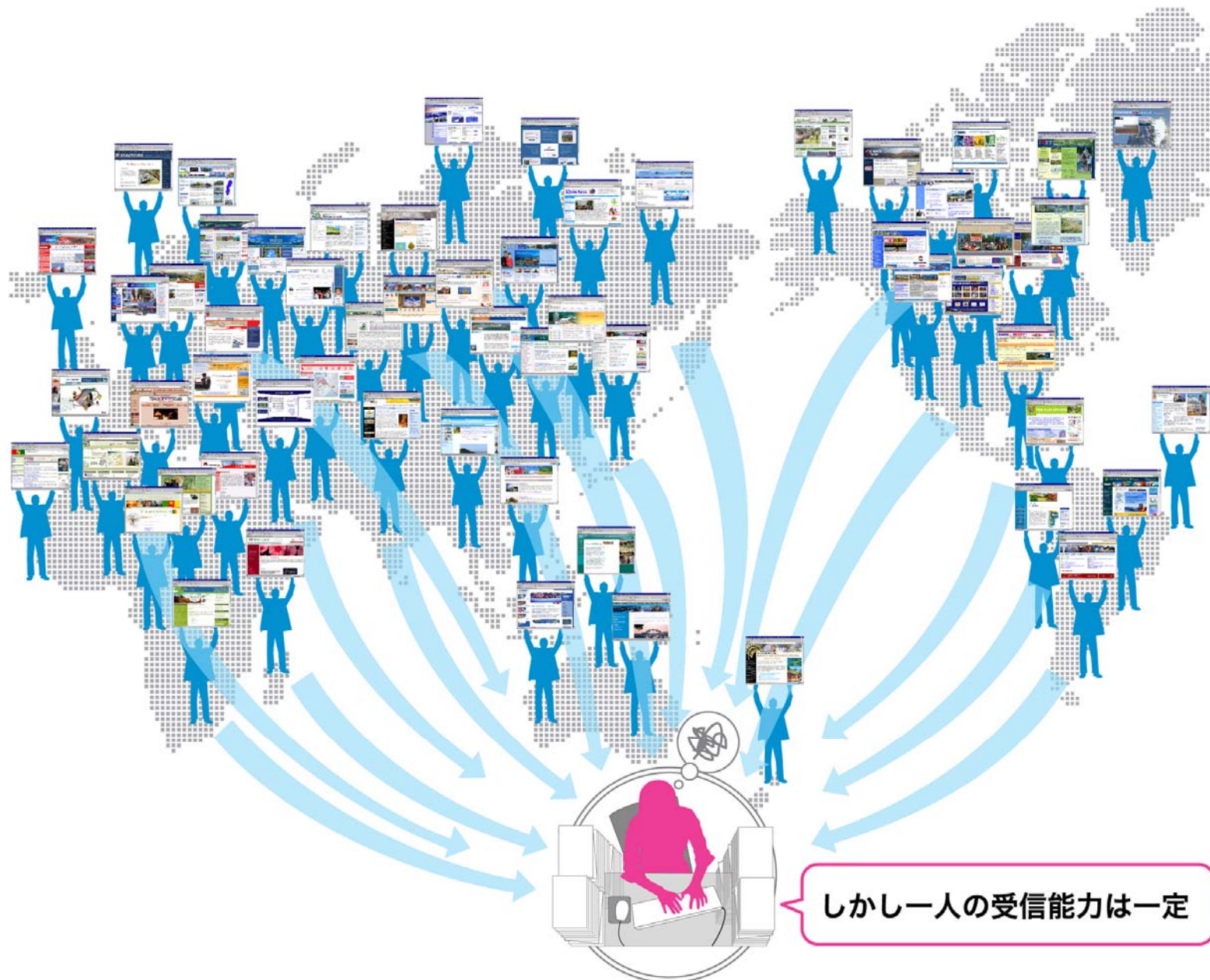
## 支援班

安達淳(NII)

# 研究の全体像



# 全人類が情報発信



# 「探す」ことに明け暮れる



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 . . .**

**複雑な質問の50%は答えが見つけれない  
 75%の人々が家族よりもWebに尋ねる**



## 難しいサーチへの挑戦



# 難しいサーチへの挑戦

● Non-1-sec response SEARCH

● 深い自然言語処理による検索

# ディープNLPサーチエンジン基盤

## TSUBAKI

### (皆でいじれるコンポーザブル型爆発サーチエンジン)



- 日本語ウェブ文書5,000万件を検索対象とした**開放型検索エンジン基盤**
  - 高度ウェブ処理用標準フォーマットによりウェブ文書を管理

相澤彰子(NII)

安達淳(NII)

乾健太郎(奈良先端)

岩山真(東工大)

宇津呂武仁(筑波)

大山敬三(NII)

奥村学(東工大)

鍛冶伸裕(東大)

神門典子(NII)

清田陽司(東大)

黒橋禎夫(京大)

新里圭司(京大)

関根聡(ニューヨーク大)

徳永健伸(東工大)

鳥澤健太郎(NICT)

中川裕志(東大)

二宮崇(東大)

藤井敦(筑波)

森辰則(横国)

吉岡真治(北大)

吉田稔(東大)

喜連川優(東大)

# クエリ: 高齢化社会で成長が見込める市場

老齡 老年 世の中 世間 発展 見込む+可能 見込む マーケット 11

Google 高齢化社会で成長が見込める市場

ウェブ 高齢化社会で成長が見込める市場の検索結果約 3,100 件中 1 - 50 件目 (0.34)

「高齢化」の検索結果 - 東京のとある大学生につき - 22-12  
 その中で、新興で、トップに迫る勢いで成長中のルビが日本の製薬会社を買収した。彼らはイン  
 ド社会で探  
 めなくな  
 d.hatena.ne  
 キャンシュ

×: 「高齢化」「市場」がブログの別  
記事に存在(クエリと関係なし)

[PDF] 金融市場展望と投資戦略 金融市場展望と投資戦略  
 ファイルタイプ: PDF/Adobe Acrobat... HTMLバージョン  
 際金融市場  
 かけて世 界  
 口減少が早

×: 日本の投資家が今後投資す  
べき分野に関するレポート

www.kokusai-amr.co.jp/reports/yamada/2004/0406.pdf - 関連 ページ - メモをとる

世界投資探検Vol.31【中国株のポテンシャル】> 成長期の日本株が ...  
 このようなブ  
 成長が見込  
 資源価格の  
 archive.mag  
 キャンシュ - 関連ページ - メモをとる

×: 金融市場に関するメールマガ  
ジンの記事(高齢化関係なし)

Welcome to IVIS INC. - よこそアドバイスへ  
 少子高齢化  
 本社会はど  
 争の中、技

×: 入社式での社長挨拶(クエリと  
関係なし)

情報爆発プロジェクト 検索エンジン基盤 Tsubaki Mozilla Firefox

高齢化社会で成長が見込める市場

で検索された文書 63 件のうち 1 - 10 件目

○: シニア向けビジネスのサポート・コン  
サルティング

http://b-model.jp/blog/cat3/post\_196.php

2 CRI(ちばさん総合研究所)千葉県経済調査研究情報-高齢社会...  
 id=044689778, score=66.7398, 類似-関連ページを表示 (1 件)

介護 市場は 高齢者 の一つの需要に過ぎず、高齢者 市場は介護 市場 以外の分野で多様化し、市場 規模も拡大するとみる向きが多い(65歳以上 市場 規  
 模: 2000年34.6兆円→ 2015年60.1兆円[ +73.7%]、電通推計)。... 高齢者 市場は、その潜在的な規模からみても、飛躍的に拡大...

○: 高齢者市場と高齢社会に対応する  
産業の動向調査

http://www.crinet.co.jp/contents/data/research/letter/175.html

3 高齢化社会の起爆剤 リバース・モーゲージ制度(1)  
 id=025151342, score=65.9271

?日本総合研究所の研究報告では、「リバース・モーゲージが20年後の2020年にア  
 国内市場で普及すると、本格的な少子高齢化 社会と長期的低経済 成長 率の持続を考えると、同制度の普及促進は経済政策の観点からも不可欠といえる。

○: 高齢者の持つ不動産を担保に融資を  
行う市場(リバース・モーゲージ)

○: 高齢化社会で成長が期待される  
医薬品業界・医療介護業界

○: 癒しを求める生活者意識の  
変化で急成長するペット市場

市場=マーケット

... 後者は高齢化・長寿化社会において  
大きな成長が見込める市場に対する  
サービス提供ということになる。...

今後も安定した成長が見込める魅力あ  
るマーケット。...

Google 高齢化社会 成長 見込める 市場

ウェブ 高齢化社会 成長 見込める 市場の検索結果約 604,000 件中 1 - 50

[PDF] 本格拡大見込める東欧圏  
 ファイルタイプ: PDF/Adobe Acrobat  
 金融部門の安定は、企業成長にとっては、不可欠の前提条件である。しかし、東欧国の株式市場  
 や、金融仲介機能。そして、流動性の程度等を見ると、EU... 高齢化の進展による将来の社会保  
 障費などの負担増加の懸念は、東欧諸国も直面してい  
 www.dir.co.jp/research/report/capital-mkt/capmkt/06  
 関連ページ - メモをとる

×: 東欧諸国経済の拡大の可能性に  
ついて

大和総研/高齢化 - 22.04  
 ... 人口構造変化の地域間格差とその影響。経営戦略研究所 主任研究員 中里 幸聖。  
 [2007.02.21] 回帰時代の地方移住は見込めるか、資本市場調査部 鈴木 準 ... 資本市場 調査部  
 尾野 功一。[2004.12.30] 超高齢化社会の到来。資本市場調査部 鈴木 準 ...  
 www.dir.co.jp/keyword/08\_aging.html - 34k - F  
 www.dir.co.jpからの検索結果 >

×: 高齢化社会の解説

○: 高齢者の持つ不動産を担保に融資を  
行う市場(リバース・モーゲージ)

「新世代」住宅市場の将来展望 - 23-15  
 長期的には、景気先行きの不透明感、建て替え比率の  
 公庫の融資金利変更等が影響し、右肩上がりの市場  
 宅業界では、高齢化社会と住宅金融公庫特別融資の開始を背景として高師...

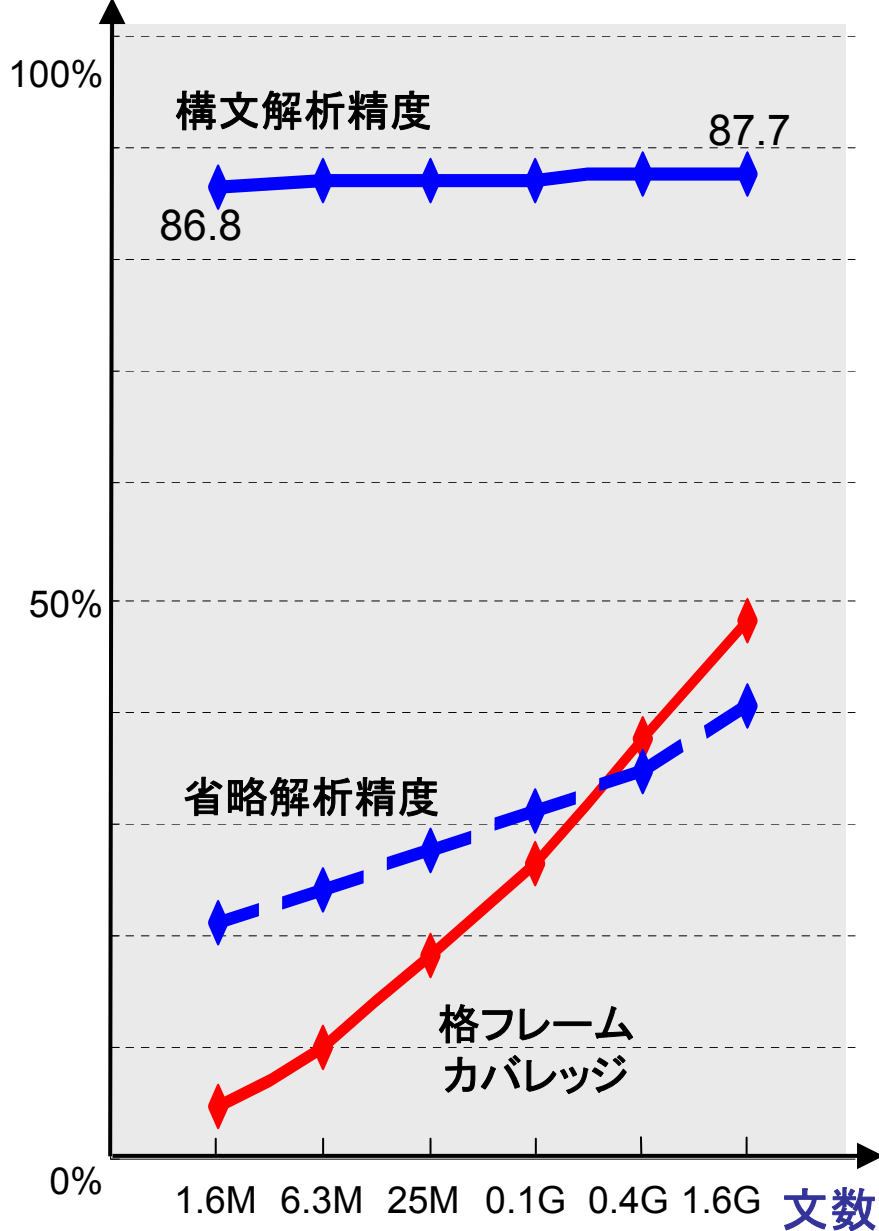
×: 住宅ビジネスの将来展望(高齢化関  
係なし)

## 格フレーム

見込める・見込まれる		成長	
ガ	増加, 拡大, 成長, 需要, ...	ガ	市場, 会社, マーケッ ト, コマース, ...
ニ	範囲, 人, 建設, 事業, 企業, イベント, 市場, 回収, ...	デ	スピード, ペース, 勢 い, 流入, 増加, ...
デ	影響, 産業, 増加, 規模, 社 会, ...	ニ	倍, 範囲, 次々, ...
	:		:

# 更なる検索の高精度化

カバレッジ・精度



## 日本語は省略が極めて多い 省略解析による検索の高度化

検索: 高齢化社会で成長が見込める市場

15 '99ペット関連市場マーケティング総覧

ペット関連市場は、新たな成長が見込まれている。この成長を支えるのは、高齢化社会の到来であり、...

16 プレスリリース | 明治乳業企業情報サイト

流動食市場は最近5年間で平均年率7-8% (当社推定)にて成長しています。また今後も、高齢化社会の拡大および医療における栄養管理の重要性の理解浸透、医療費削減施策等により、一層の成長性が見込まれます。

格フレーム

見込める・見込まれる

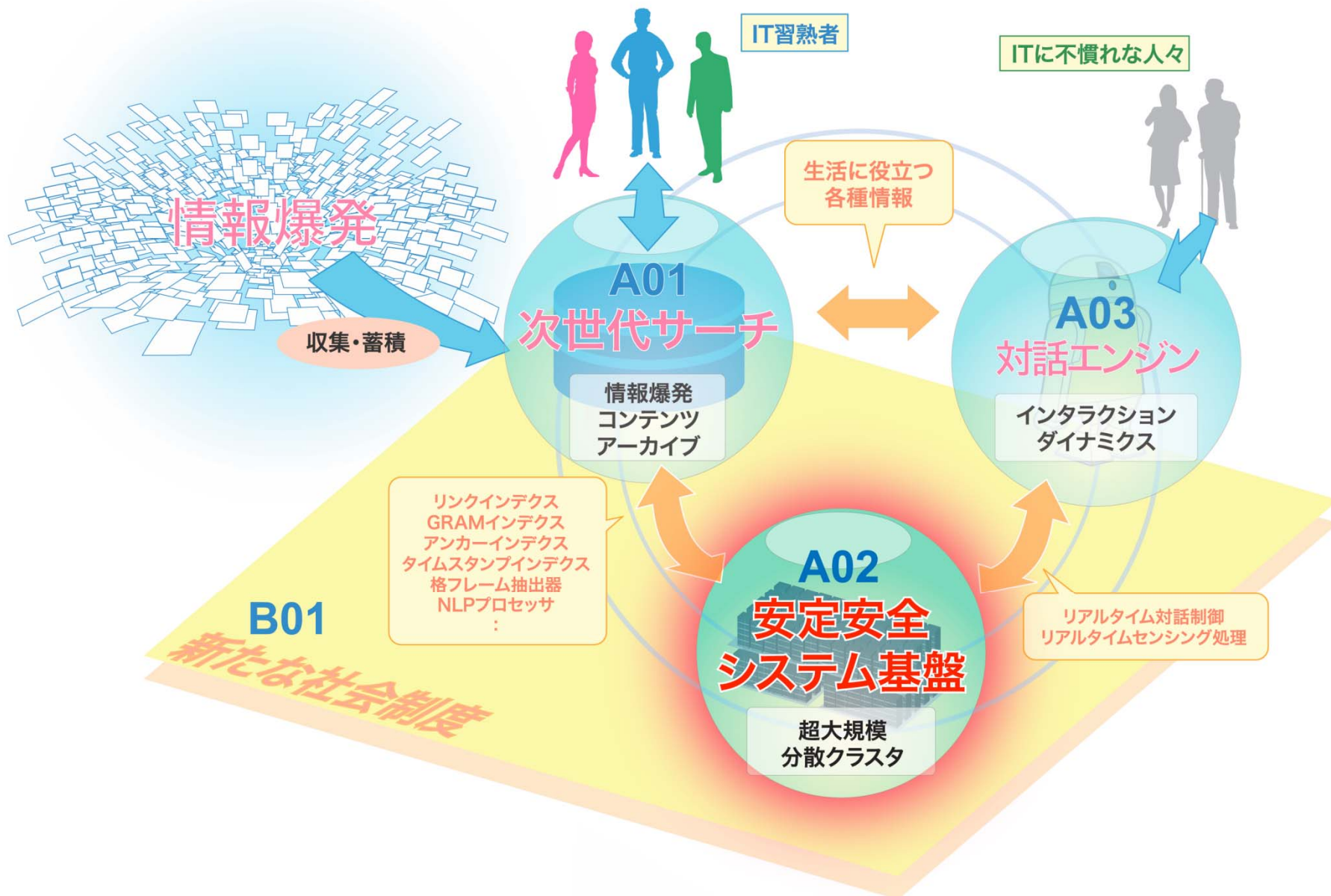
ガ	増加, 拡大, 成長, 需要, 効果 ...
ニ	範囲, 人, 建設, 事業, 企業, イベント, 市場, 回収, ...
デ	影響, 産業, 増加, 規模, 社会, ...

# 研究の推進体制

特定領域研究として、柱間で研究資源を共有し、効果的に研究を推進する体制を実現

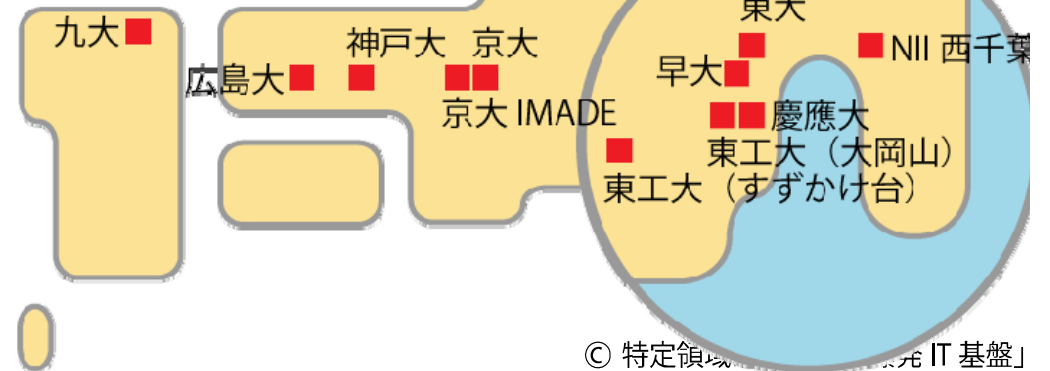
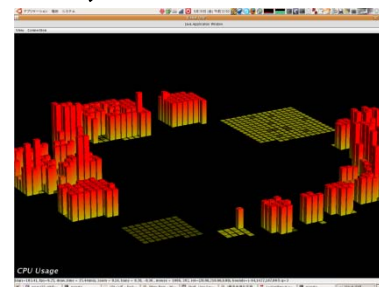
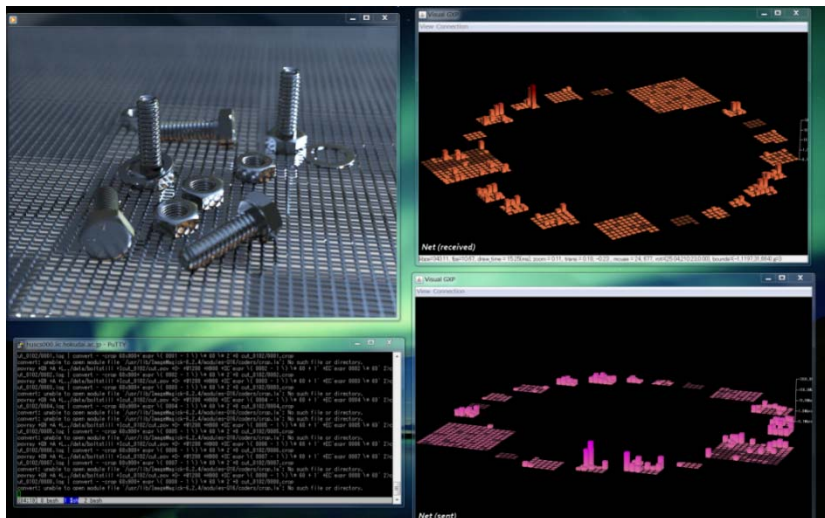
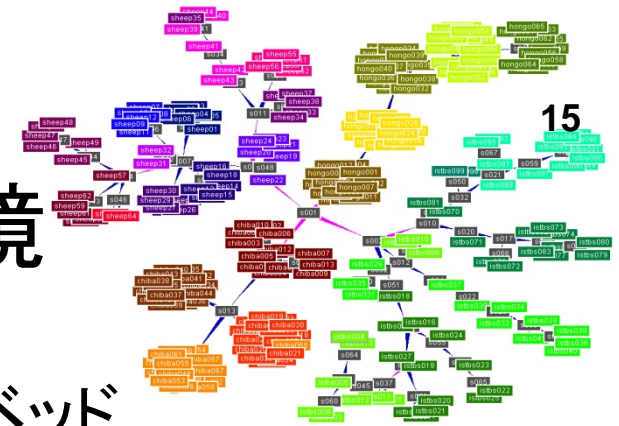


# 研究の全体像

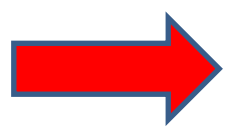
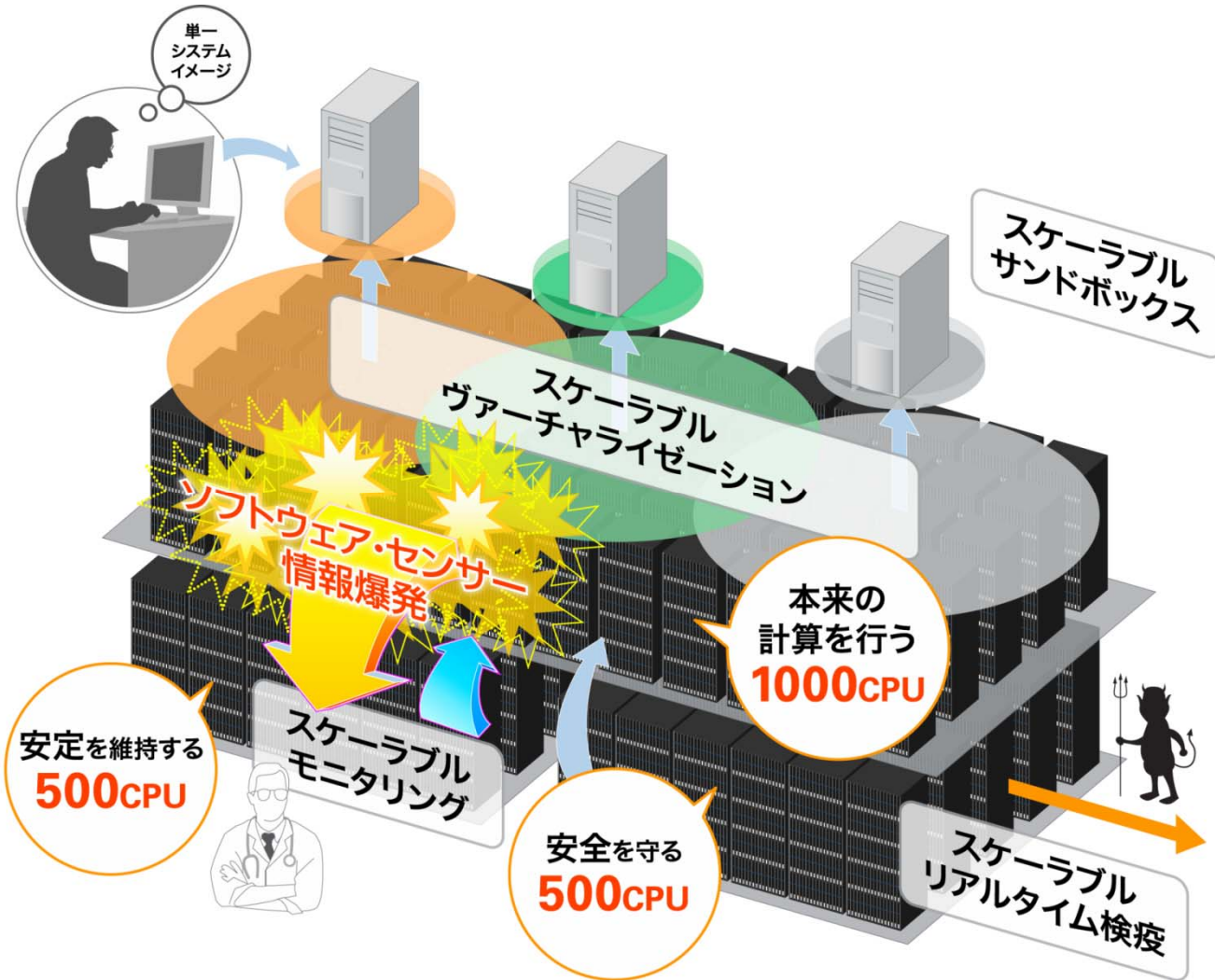


# InTrigger:15拠点, >1600 CPUコアの実験環境

- 約58グループが利用
  - システム研究者の多拠点並列処理テストベッド
  - 情報処理基盤研究者の計算資源
- GXP並列シェル, GXP make, GXP MapReduce
- Gfarm広域ファイルシステム, GMount「即席」ファイルシステム
- 小規模クラスタ, スパコン, 分散環境まで, 共通の記述・操作で並列実行を可能に!



# 安定・安全な大規模システム基盤



**MPICH2** (定評あるオープンソースMPI実装)における  
10年来のバグ(多ノードで非決定的にデッドロックダウン)の同定  
(1テラバイトログマイニング) → 国際コミュニティへの貢献



**New IT Infrastructure for the Information Explosion Era**

Research Project funded by MEXT Grant-in-Aid for  
Scientific Research in Priority Areas (FY 2005-2010)

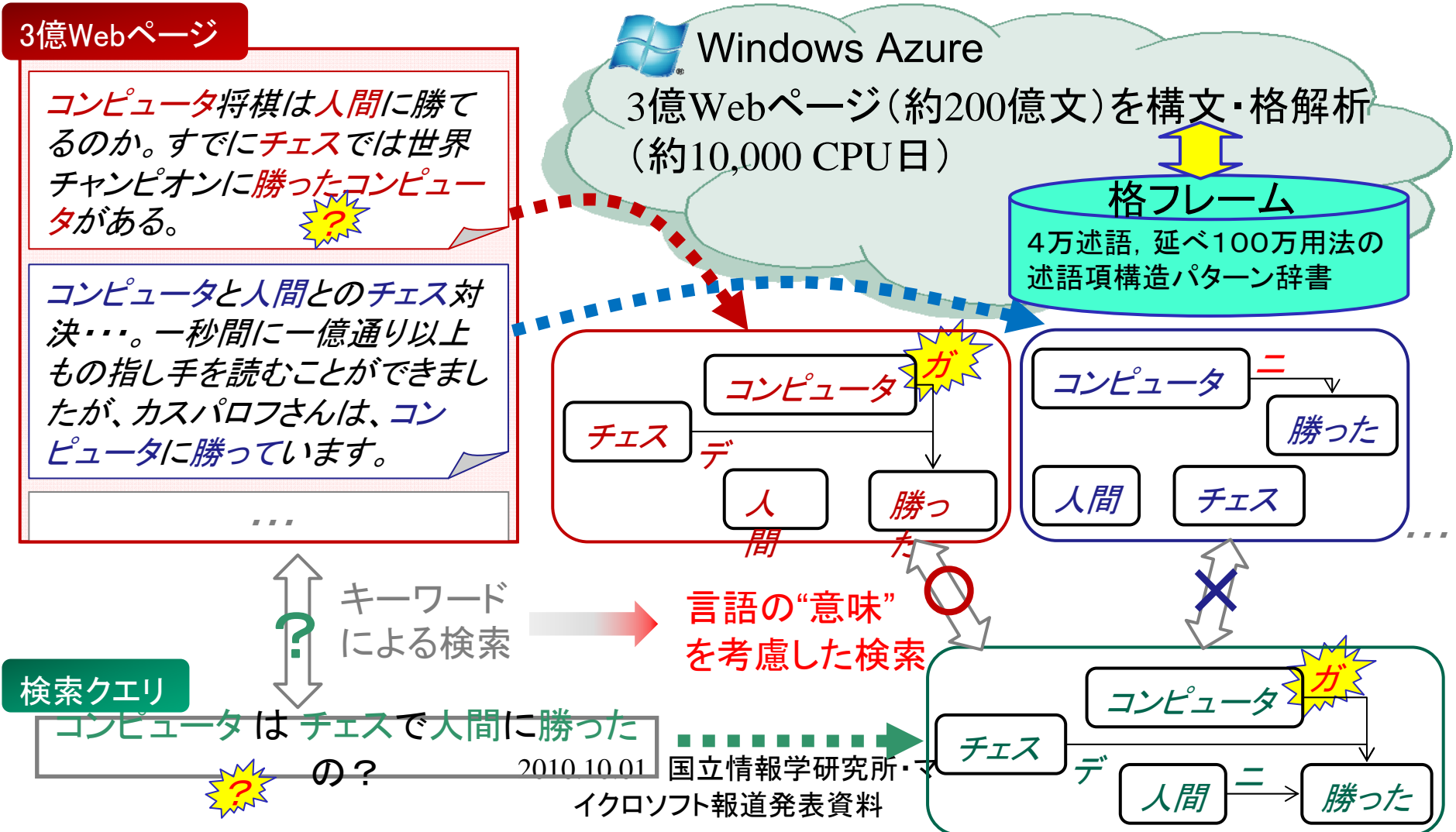
*info*-plosion

# Azureクラウドとの連携

# 超大規模Webテキストの構文・格解析

—言語構造を考慮する検索エンジンTSUBAKIの深化へ向けて—

黒橋禎夫, 河原大輔, 柴田知秀, 新里圭司, 笹野遼平(京大)



# テキストから書き手の「心」を推測する

- テキストから出来事や意見を抽出し、
- その出来事や意見に対する書き手の心的態度や事実性を推定。
- その情報を索引化し、検索可能に。
- 社会で広く共有できる「知の宝庫」へ。

次世代検索サービス  
 - 意見・事実検索  
 - マーケティング、データマイニング、社会分析



TSUBAKI

Azureクラウド



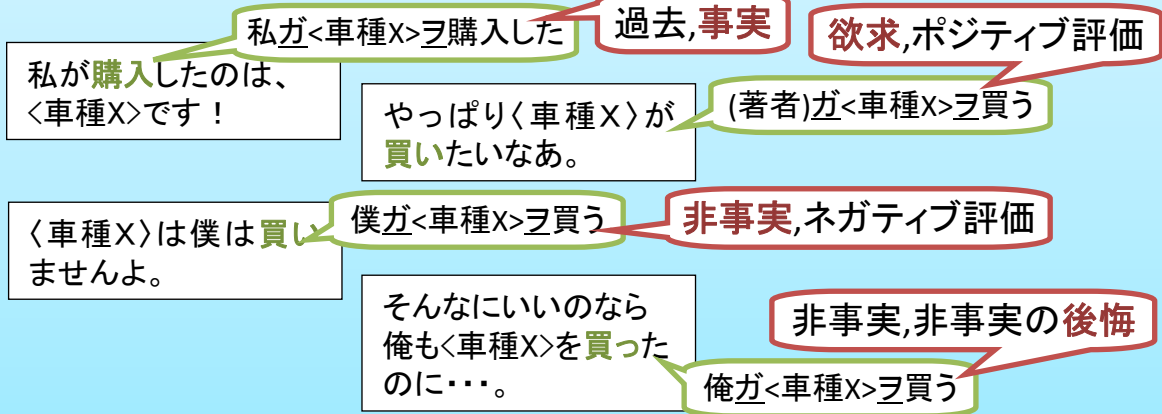
3億の日本語Web文書を高速に解析

格解析 (ガヨニデ推定)

拡張モダリティ解析 (心的態度・事実性推定)

さまざまなピックに関する多様な出来事や意見が、書き手の態度(宣言, 欲求, 意志, 推量, 高評価, 後悔など)とともに記述されている

- やっぱり<車種X>が買いたいなあ。
- 昨日、夫が<車種X>を買いました。
- 私が購入したのは、<車種X>です！
- そんなにいいのなら俺も<車種X>を買ったのに…。
- <車種X>は僕は買いませんよ。
- ぜひ<車種X>をご購入ください。

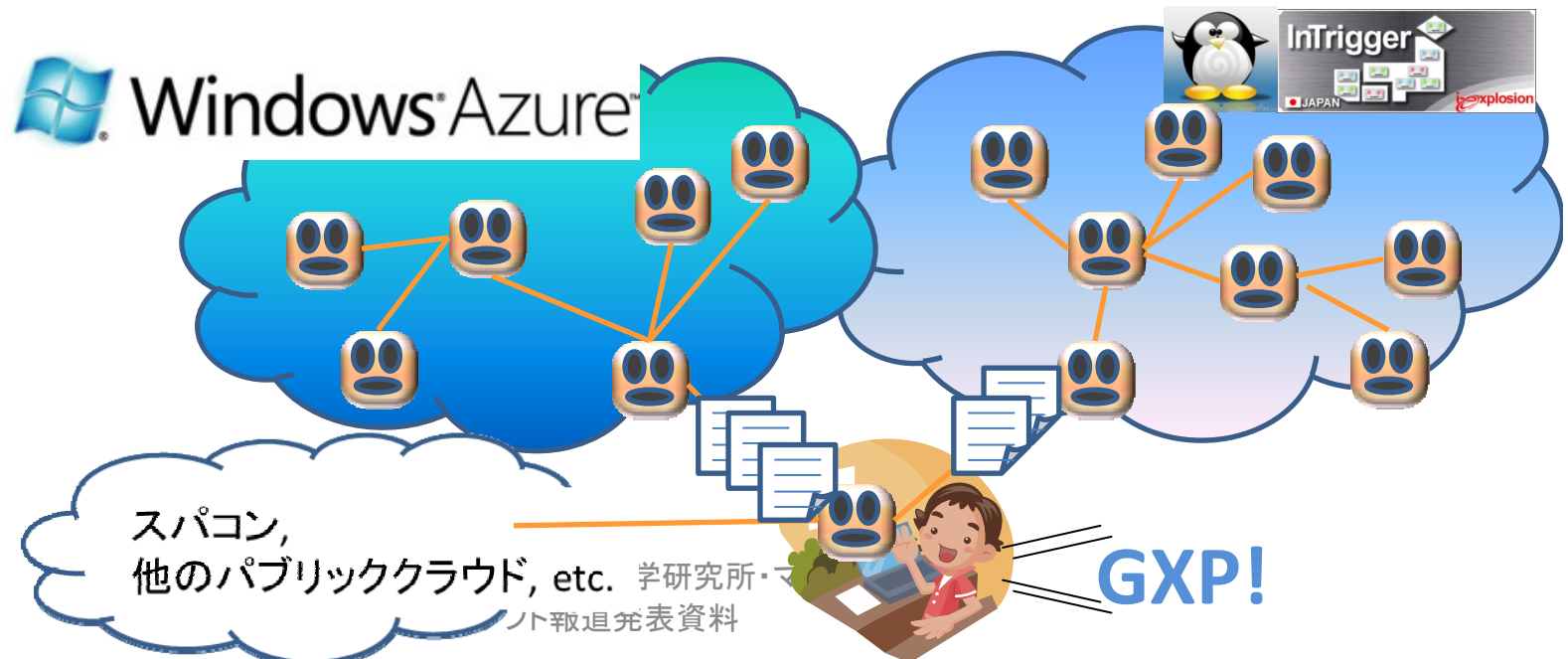


⋮

# ヘテロクラウド連携を可能にする 並列スクリプティング・ワークフロー環境

- **Windows Azure**, Linux クラスタ, 他のパブリッククラウド, スパコンなど, 多種多様な計算機資源をシームレスに統合利用できる**並列スクリプティング・ワークフロー環境**を開発する
- 情報爆発プロジェクトの成果をベースに開発
  - InTriggerプラットフォーム
  - GXP並列ワークフロー環境

情報爆発InTrigger  
プラットフォーム

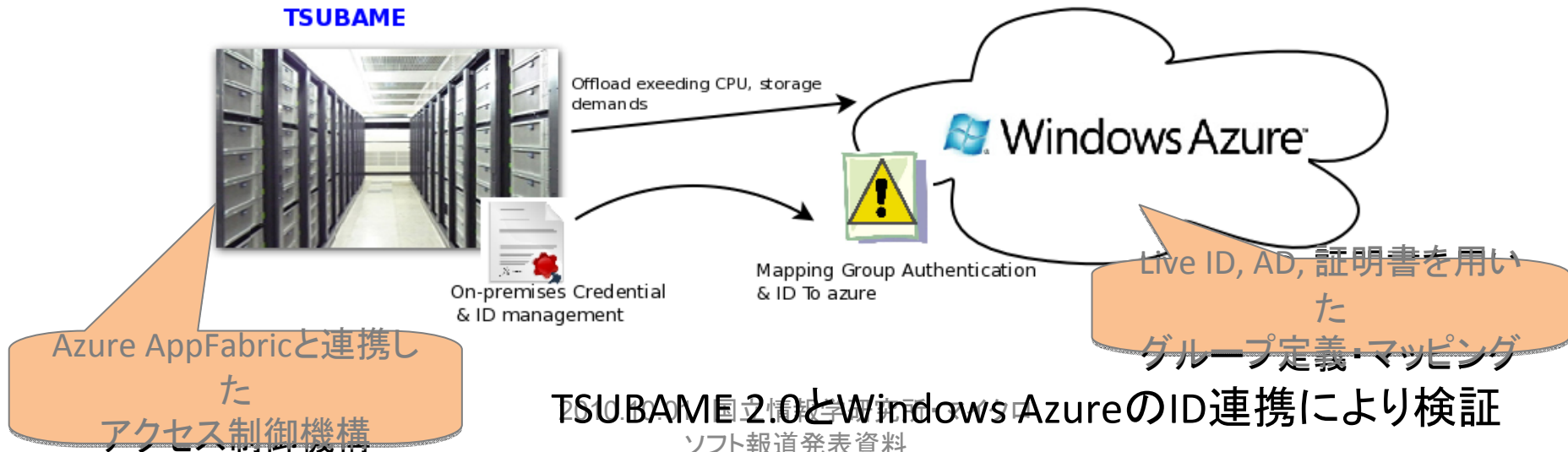


# スパコン・クラウド連携 認証・認可技術

研究代表者: 松岡 聡 (東京工業大学/国立情報学研究所)

研究チーム構成メンバ: 丸山 直也, Mohamed Amin JABRI (東京工業大学)

- e-サイエンス、科学技術計算によるクラウド利用
  - 突発的なCPU,ストレージリソース要求の受け皿
  - グループシェアされている学術研究データの永続保持、災害回避
- ➡ オンプレミス(スパコン)のセキュリティポリシーを満たしつつ、作業環境をクラウドまで拡張することが必要
- スパコン・クラウド連携のための課題と研究項目

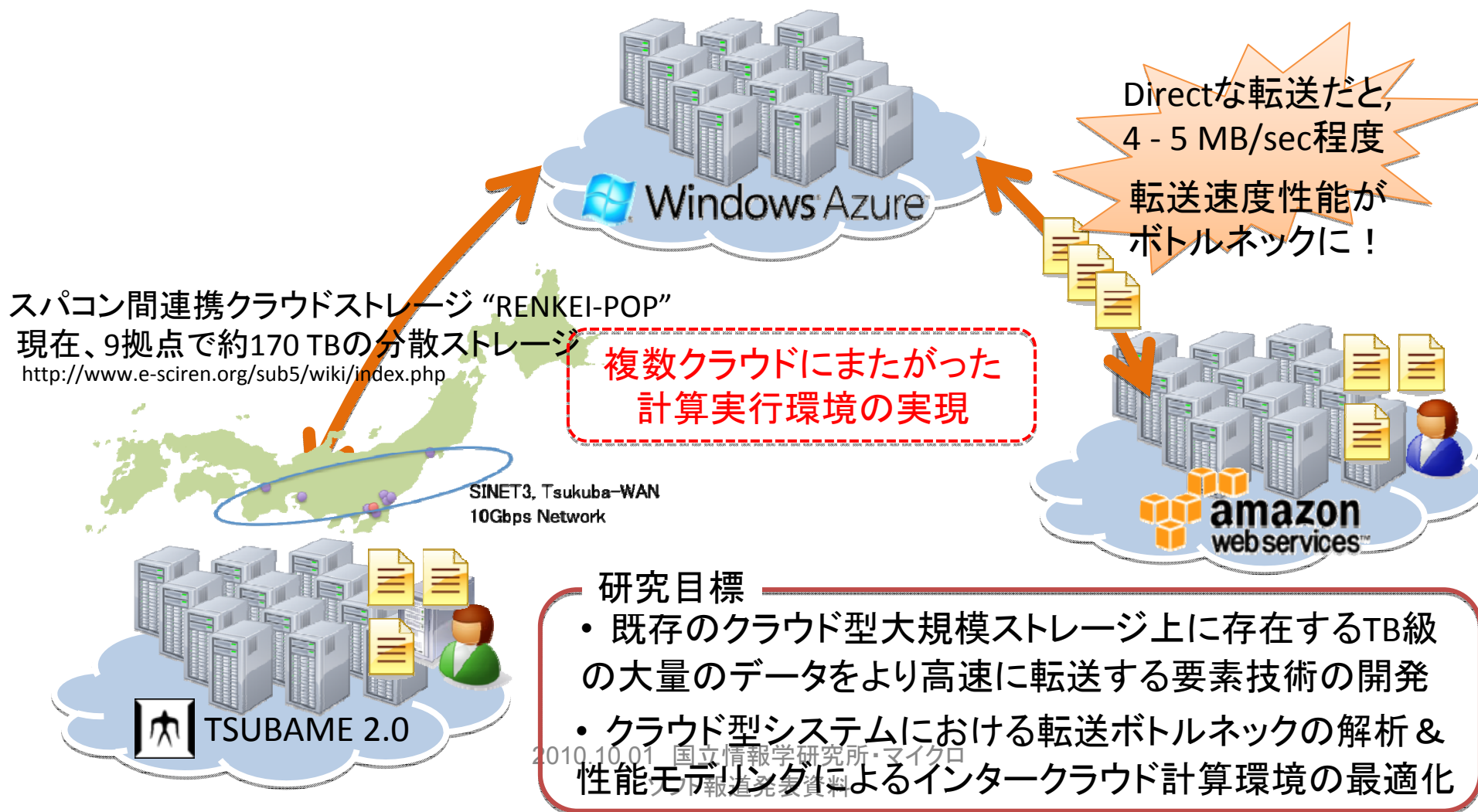


# インタークラウド間の大規模データ転送

Windows Azure Platform, Amazon Web Services, TSUBAME 間でのe-Scienceデータの共有を目指して

研究代表者: 松岡 聡 (東京工業大学/国立情報学研究所)

研究チーム構成メンバー: 丸山 直也, 滝澤 真一郎, 千葉 立寛, 佐藤 賢斗 (東京工業大学)



# 大規模クラウド連携

研究代表者: 松岡 聡 (東京工業大学/国立情報学研究所)

研究チーム構成メンバ: 丸山 直也, Sumeth Lerthirunwong (東京工業大学)

- 世界中の様々なクラウドを有機的に連携
- 例えば・・・データ移動コスト大 → データソースに近いクラウドを活用、遠くのクラウドにも混雑度に応じて動的にデータ&計算を移動

動的に変化する環境への対応は？

クラウド間でのデータのコンシステンシは？

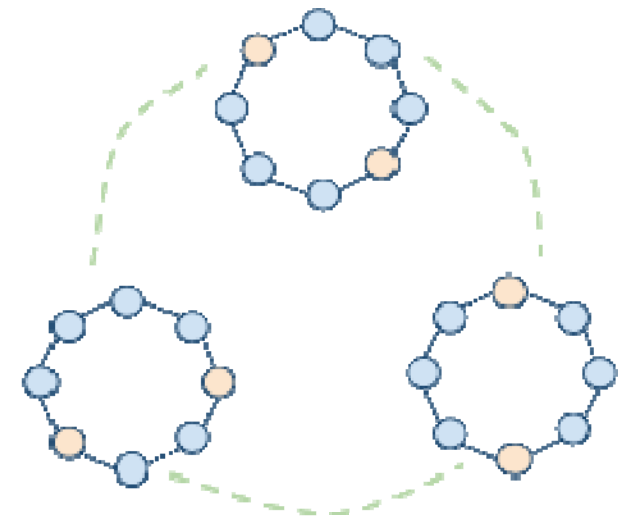


## 「マルチリング構造化オーバーレイ」

リング構造ネットワークを拡張  
負荷分散 & 局所性の活用  
高効率

### TSUBAME2とAzureの連携により実証

2010.10.01 国立情報学研究所  
ソフト報道発表資料



TSUBAME 2



Windows Azure

# Windows Azureによる超大容量データの解析 ～日常生活を支援するCrowdsourcing～

Huh?

基本料金  
電力量料金  
上記料金  
・1段料金  
・2段料金  
・燃料費調整額

➤外国人が日本語を辞書で調べる事は困難  
➤画像処理で抽出可能な情報は一部のみ

Snap!  
Tell me what it says ...

