

グラフを辿って関係発掘 CSM (Context Structure Matching)

日本電信電話株式会社 NTT情報流通プラットフォーム研究所
佐藤 宏之 飯塚 京士 イコプラムディオノ 村山 隆彦

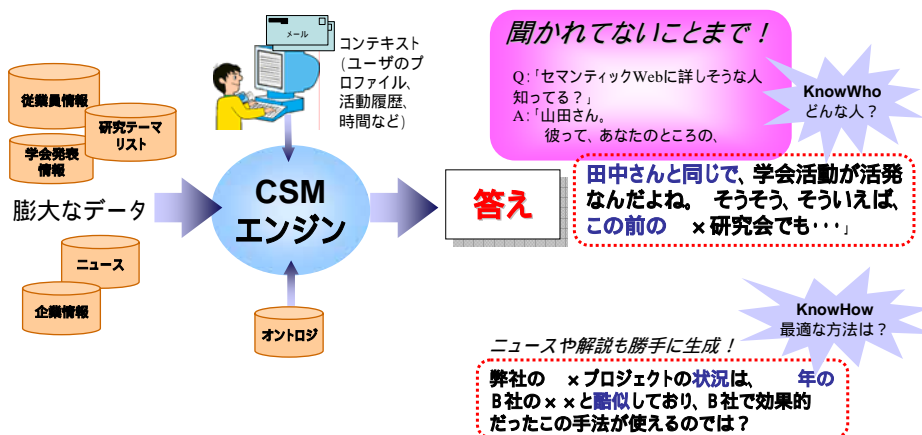
NTTソフトウェア株式会社
妹尾 正身



(c)2006 NTT Information Sharing Platform Laboratories and NTT Software Corporation. All rights reserved.

CSMの狙い

- 膨大な情報の中から「特徴的かつ有用な」「状況・背景にあった」「**答え**」を見つけたい



(c)2006 NTT Information Sharing Platform Laboratories and NTT Software Corporation. All rights reserved.

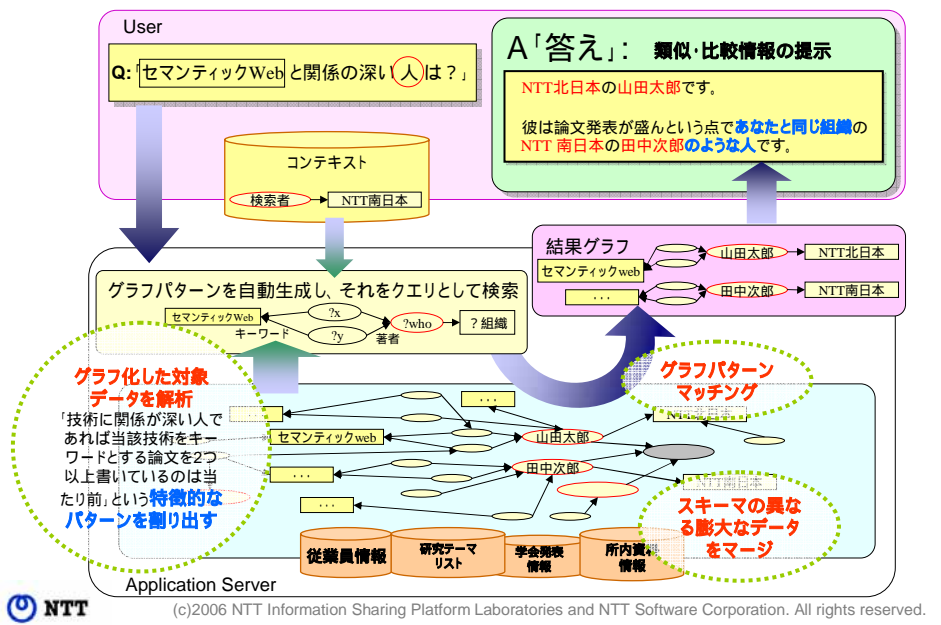
アプローチ

- 情報の再構成による新たな知識発見
 - ヘテロな構造を持つ膨大なデータをグラフ構造に変換してマージ
 - RDFのデータモデルを利用
 - クエリグラフパターンを自動生成
 - マージされたデータのグラフ構造の解析
 - コンテキスト(ユーザプロフィールや検索履歴)の利用



(c)2006 NTT Information Sharing Platform Laboratories and NTT Software Corporation. All rights reserved.

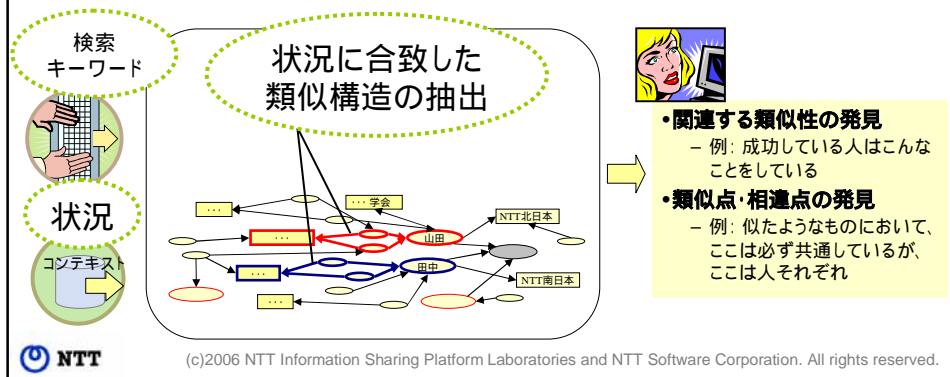
CSMの動作概要



(c)2006 NTT Information Sharing Platform Laboratories and NTT Software Corporation. All rights reserved.

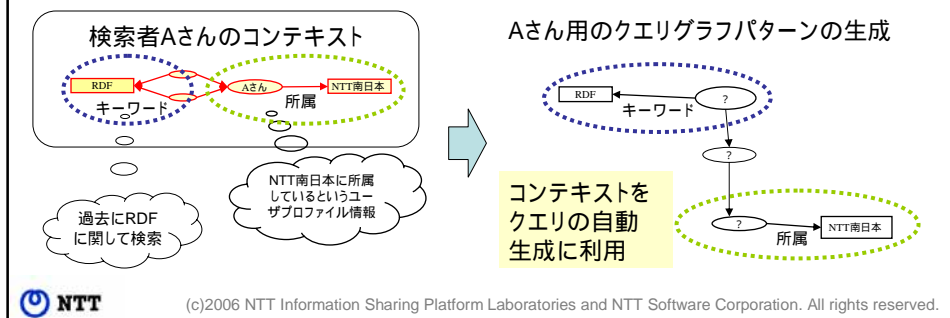
技術のポイント(1)

- グラフパターンマッチングによる知識の発見
 - グラフ化してマージすることでスキーマの違いを吸収
 - グラフの頻出構造の解析などから、知識表現における特徴的な類似性を抽出するためのクエリグラフパターンを自動生成



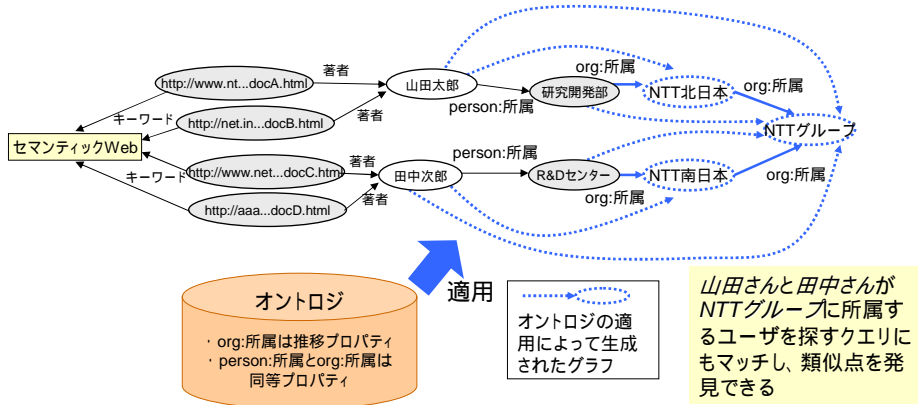
技術のポイント(2)

- コンテキストの利用
 - ユーザコンテキスト(プロフィールや活動履歴)もRDF化
 - グラフパターンマッチングの際に
 - 探索範囲の絞り込み
 - ユーザの興味に基づいたクエリパターンの自動拡張



技術のポイント(3)

- オントロジを利用したグラフ構造の変形による柔軟なマッチング



(c)2006 NTT Information Sharing Platform Laboratories and NTT Software Corporation. All rights reserved.

CSM検索結果例(1)

検索キーワード: セマンティックWeb

【答え】
「セマンティックWeb」と関係の深い人は以下です

山田太郎
田中次郎

【理由や背景】

| ドキュメントの著者 | ドキュメント | ドキュメントのキーワード | 著者の所属 |
|-----------|----------------------------|---------------------------|-------------------|
| 山田太郎 | http://www.nt...docA.html | セマンティックWeb RDF | NTTグループ NTT北日本 |
| | http://net.in...docB.html | | |
| 田中次郎 | http://www.net...docC.html | セマンティックWeb ソーシャルネットワーク | NTTグループ NTT南日本 |
| | http://aaa...docD.html | | |

類似点: 2人ともセマンティックWebをキーワードとするドキュメントを2つ以上書いた

類似点: 2人ともNTTグループ、田中次郎は検索者と同じ組織に属している

比較点: セマンティックWeb以外のキーワードは異なる

比較点: 山田太郎は検索者とは異なる組織に属している

クエリグラフパターンとしてプロパティ(アークの意味と構造)のみを利用することで、セマンティックWebと関係が深い、それぞれの人はこのような点で異なるといった結果が出せる



(c)2006 NTT Information Sharing Platform Laboratories and NTT Software Corporation. All rights reserved.

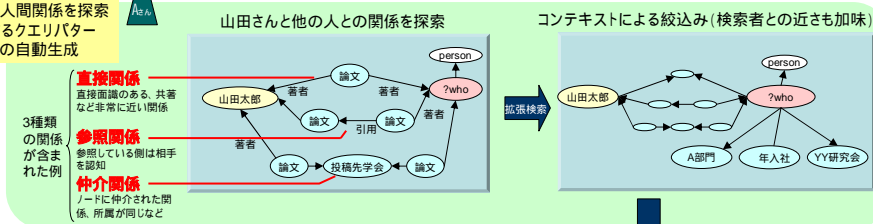
CSM検索結果例(2)

| ドキュメントの著者 | ... | 著者の所属 |
|-----------|-----|-------------------|
| 山田太郎 | ... | NTTグループ NTT北日本 |

検索結果の山田さんは検索者Aさんと別の所属だった...

グラフの解析から、**複雑かつ特徴的な人間関係**を探索するクエリパターンの自動生成

山田さんについて教えてくれそうな人が身近にいるだろうか？



【結果と理由】

| セマンティックWebと関係が深い人 | 山田太郎と関係が深い人 | あなた(検索者)との近さ |
|-------------------|--------------|-----------------------------|
| 山田太郎 | 中村三郎 鈴木四郎 | • 同期 • 同じ所属 • 同じYY研究会 |



(c)2006 NTT Information Sharing Platform Laboratories and NTT Software Corporation. All rights reserved.

おわりに

- 今後さらに有用な知識を抽出できるようにクエリグラフパターン生成のアルゴリズムを検討
- オフィスのさまざまなデータに適用し、異なる種類のデータベースを横断して「類似する人」や「比較対象となる技術」を検索する実験を実施中



(c)2006 NTT Information Sharing Platform Laboratories and NTT Software Corporation. All rights reserved.