

LinkedDataアプリの構築経験から感じた LODの課題

セマンティックWebコンファレンス2011
パネルディスカッション「日本にも広がるか? LOD」資料

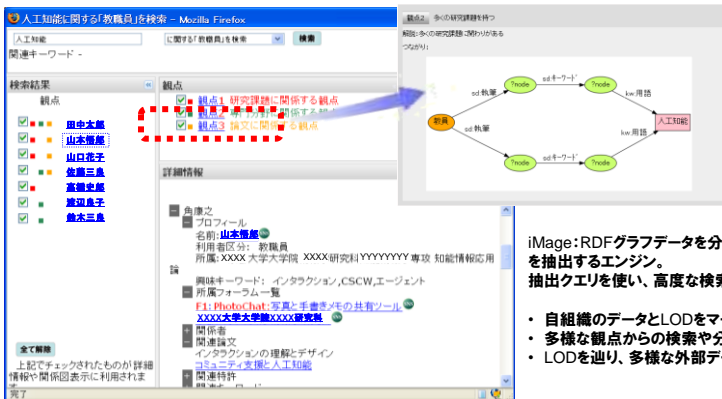
飯塚 京士

© 2011 NTT Information Sharing Platform Laboratories

LODから得られる恩恵

- LODが普及すれば、自分のデータとLODをリンクさせて高度な分析が行える
 - 様々なデータソースを基に、多角的な分析
 - 容易に集合知を活用

自組織のデータとLODを使用したアプリケーション例 iImage (NTT研究所の技術)



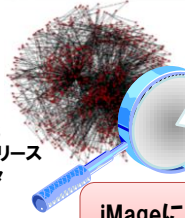
The screenshot shows a web browser window with a search interface. On the left, there's a search bar and a list of search results with checkboxes. A red dashed box highlights three specific results. A blue arrow points from these results to a graph visualization on the right. The graph shows nodes connected by edges, representing relationships between data points. Below the graph, there's a detailed view of a selected item, including its name, affiliation, and keywords. To the right of the graph, there's a text box describing the iImage engine and its capabilities.

iImage: RDFグラフデータを分析し、特徴的なクエリを抽出するエンジン。
抽出クエリを使い、高度な検索アプリを構築できる。

- 自組織のデータとLODをマージした分析が可能
- 多様な観点からの検索や分類が可能
- LODを辿り、多様な外部データへのリンクが可能

異なる情報源のRDFデータを統合してできたグラフ構造

- 例)
- 研究者DB
 - ニュースリリース
 - SNSデータ
 - ...

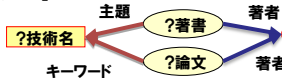


iMageによる解析

特定ノード間の頻出パターンのマイニングなど

データの中から特徴的なパターンを発見する

- 例)
- 「技術と人とのパターン」



第1の観点:
ある技術に関する著作が多い

技術に詳しい人を探すときに使えるパターン

iMageを使ったアプリ構築経験から感じられた、LOD導入の障壁

公開されているLODを利用するとき

- データの特徴や品質等に関する情報が見当たらず、利用しにくい。
 - データの信頼性は？
 - 名寄せが行われたクリーンなデータか？
 - リソースURIの命名規則は？
 - スキーマ定義情報はどこにある？
 - 他データのスキーマと整合性はとれている？
 - 意味が似たデータが複数ある場合の違いは？

自分のデータをLinked Data化するとき

- データを表現するボキャブラリはどうする？
 - 自分のデータの意味と、汎用ボキャブラリ(FOAFやDoblin Coreなど)の意味が微妙に違い、悩ましい。
 - 汎用ボキャブラリを使うメリット: LODとの接続性が良い
 - 独自ボキャブラリを使うメリット: 微妙な意味の違いを自由に表現できる。
- 名寄せはどうする？
 - 適切にリンクを形成するように、名寄せは必要不可欠。
 - LODの既存リソースと名寄せする必要があり、場合によってはこの作業が大変
 - そもそも、自分が持つデータ内の名寄せが大変

- **LinkedData化のガイドラインの整備**
 - LinkedData化の事例集
 - リソース化の方針（名寄せ、既存LODへのリンク等）
 - 汎用/独自ポキャブラリ利用ノウハウ

- **LODの利用を促すための環境整備**
 - 日本語LOD用の簡易検索システム
 - データの使用感を試す場
 - リソース一覧などの取得など
 - 既存LODデータのスキーマ情報、ポキャブラリ情報の提供