

ISWC2016参加報告 ～Linked Data周辺～

江上周作¹, 川村隆浩^{1,2}

¹電気通信大学大学院情報理工学研究科

²科学技術振興機構

報告内容

- 以下のセッションからLinked Data周辺を報告
 - Linked Data
 - Linked Data Measurement
 - Knowledge Graph
 - Linked Data for Information Extraction (W)
 - Visualization and Interaction for Ontologies and Linked Data (W)

Linked Data

- 4日目（本会議2日目）に開催
- Chair : Alasdair J G Gray
- 発表リスト
 - Conference Linked Data: the ScholarlyData project [\[Resources\]](#)
 - FOOD: Food in Open Data [\[Resources\]](#)
 - LinkGen: Multipurpose Linked Data Generator [\[Resources\]](#)
 - Clinga: Bringing Chinese Physical and Human Geography in Linked Open Data [\[Resources\]](#)

全てResources Track paperで、Linked Data生成からオントロジーの設計まで。

Conference Linked Data: The ScholarlyData Project

著者 : Andrea Giovanni Nuzzolese, Anna Lisa Gentile, Valentina Presutti, and Aldo Gangemi

- 論文、会議情報等のLinked Open Dataの提供 (SWDF)
- Semantic Web Dog Food (SWDF)^{*1} における課題の指摘
 - ガイドライン不足, 管理維持されていない語彙の使用, クラス・プロパティの誤使用, クエリをかけにくい構造, タイポ
- 上記問題の解決
 - Semantic Web Conference ontology^{*2}のリファクタリング
→Ontology Design Patterns (ODPs)^{*3}に基づいて定義
 - データ生成ワークフローのオープンソース実装→cLODg^{*4}
 - SWDFデータセットの再生成→cLODgを使用
- scholarlydata.orgで公開
 - 総トリプル数1,128,618

FOOD: Food in Open Data

著者 : Silvio Peroni, Giorgia Lodi, Luiji Asprino, Aldo Gangemi, and
Valentina Presutti

■ イタリアのe-governmentプロジェクトFOOD

- 原産地名称保護制度 (PDO) , 地理的表示保護制度 (PGI) , 伝統特産品保護制度 (TSG) のPDF文書からデータを抽出しLODにすることが目的

■ FOODオントロジーの構築

- 2人のオントロジーエンジニアが800以上の政治文書を分析
- ODPsに基づく
- 商標名, 原材料, 原産地, 化学成分, タイプなどを含む

■ LODの半自動構築

1. スクリプトによりPDF, WORDからデータ抽出しExcelに
2. 専門家による検証、結果をマージ
3. スクリプトによりXLS→CSV→RDFに変換

LinkGen: Multipurpose Linked Data Generator

著者 : Amit Krishna Joshi, Pascal Hitzler, and Guozhu Dong

- ベンチマーク用人工Linked Dataの生成器「LinkGen」の開発
 - これまでの特定の領域を対象としたものでなく、幅広い語彙に対応している
 - ガウス分布, ジップ分布に基づくデータ生成
 - ノイズデータの追加 (プロパティの値の衝突, シンタックスエラー, 誤ったdomain/range, タイプなしインスタンス)
 - リアルデータセットとのリンク付も可能
 - VoIDダンプとRDFストリームのアウトプットも可能
 - パラメータの設定が可能

Clinga: Bringing Chinese Physical and Human Geography in Linked Open Data

著者 : Wei Hu, Haoxuan Li, Zequn Sun, Xinqi Qian, Lingkun Xue, Ermei Cao, and Yuzhong Qu

■ 中国のLinked Geographical dataset: Clingaの開発

- 地理要素とリレーションを含む
- Baidu Baikeからデータを取得（中国においてはWikipediaより大規模。特に地理や人物関係はWikiの15倍）
- GeoNamesはクラス階層がなく， Baikeカテゴリは一貫性がないため新たな地理オントロジーを設計
- SVMを使ってエラーを回避
- DBpediaやGeonamesとリンク

Clinga: Bringing Chinese Physical and Human Geography in Linked Open Data

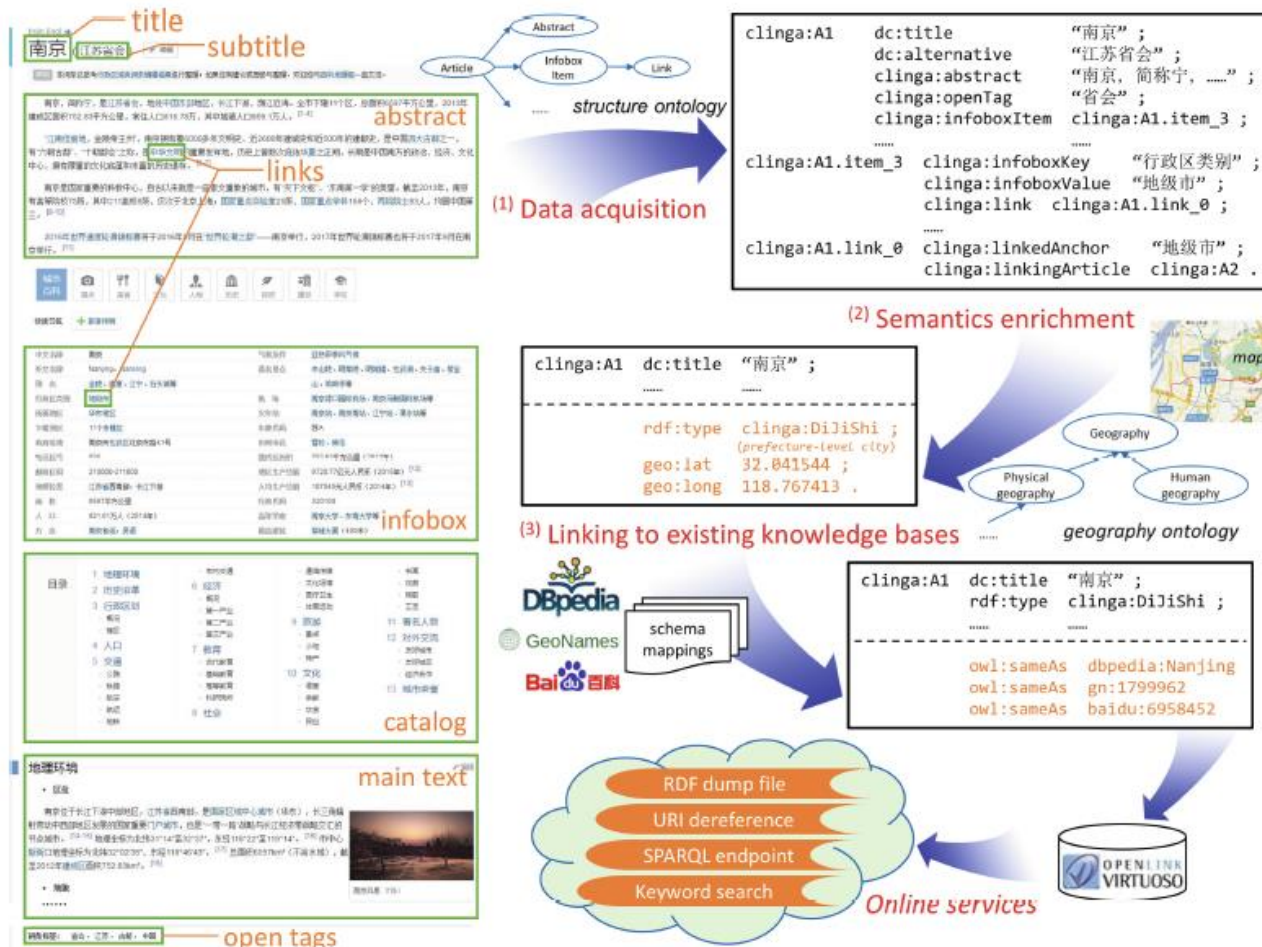


Fig. 1. Left part is a typical article in Baidu Baike, and right part is the methodological steps to develop Clinga, including: (1) data acquisition, (2) semantics enrichment, and (3) linking to existing knowledge bases.

Knowledge Graph

- 3日目（本会議1日目）に開催
- Chair : Craig Knoblock
- 発表リスト
 - VoldemortKG: Mapping schema.org and Web Entities to Linked Open Data [\[Resources\]](#)
 - Building and Exploring and Enterprise Knowledge Graph for Investment Analysis [\[Applications\]](#)
 - DeFacto – Temporal and multilingual Deep Fact Validation [\[Journal\]](#)
 - Linked Data (in low-resource) Platforms: a mapping for Constrained Application Protocol [\[Resources\]](#)

VoldemortKG: Mapping schema.org and Web Entities to Linked Open Data

著者 : Alberto Tonon, Victor Felder, Djellel Eddine Difallah, and Philippe Cudré-Mauroux

- Webページ間, LODとのエンティティリンクング
- 18億ページをクロールしアノテーションとwikiへのアンカーを含むページのソースコードを抽出
- データセットの生成にApache Spark
- ページ, 5-ple(Subject, predicate, object, page url, markup format), アンカーをParquetフォーマットで保存 (Snappyで圧縮)
- ラベルの文字列マッチングでリンク付け

VoldemortKG: Mapping schema.org and Web Entities to Linked Open Data

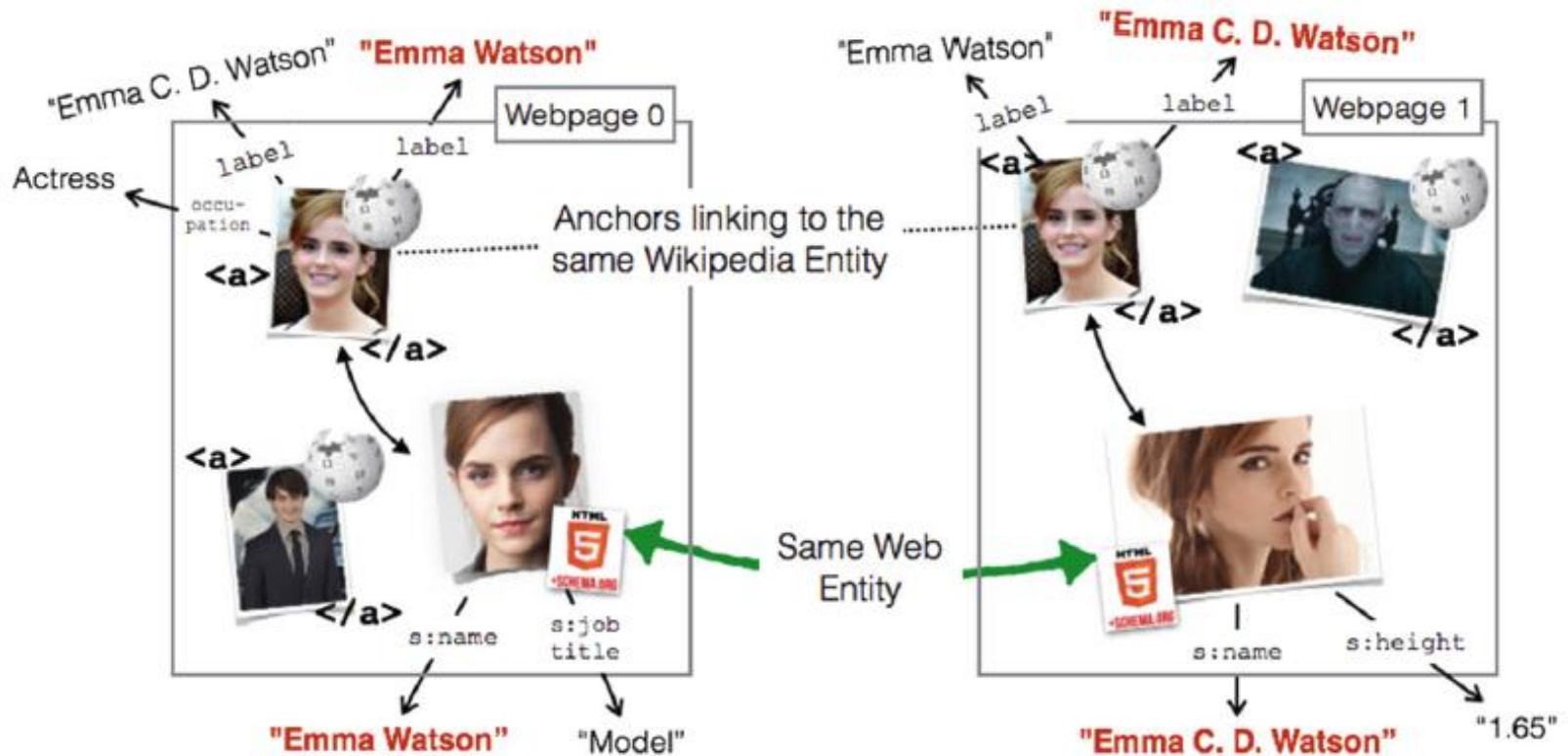


Fig. 1. Instance matching method used to build VoldemortKG. On the left-hand side, the DBpedia entry for Emma Watson is considered equivalent to a schema.org entity as its name is also a label of Emma Watson in DBpedia. On the right-hand side, a similar situation takes place for the same DBpedia entry and another Web entity. We thus conclude that all the mentioned entities refer to the same actress.

Linked Data Measurement

- 4日目（本会議2日目）に開催
- Chair : Heiko Paulheim
- 発表リスト
 - Are Names Meaningful? Quantifying Social Meaning on the Semantic Web [[Research](#)]
 - Learning to Assess Linked Data Relationships Using Genetic Programming [[Research](#)]
 - LODStats: The Data Web Census Dataset [[Resource](#)]
 - Quality Assessment for Linked Data: A Survey [[Journal](#)]

Linked Data Measurement

- Are Names Meaningful? Quantifying Social Meaning on the Semantic Web
 - 著者 : Steve de Rooij, Wouter Beek, Peter Bloem, Frank van Harmelen, and Stefan Schlobach
 - 相互情報量の尺度でIRIの意味を定量化する手法の提案
 - データセットごとに意味のある名前がどの程度含まれているか算出
- Learning to Assess Linked Data Relationships Using Genetic Programming
 - 著者 : Ilaria Tiddi, Mathieu d'Aquin, and Enrico Motta
 - 遺伝的プログラミング (GP) を使用して (トポロジカル, セマンティックの観点から) Linked Data内の結びつきの強い関係を発見する

その他ワークショップ

Visualization and Interaction for Ontologies and Linked Data (1日目開催)

■ A Visual Aide for Understanding Endpoint Data

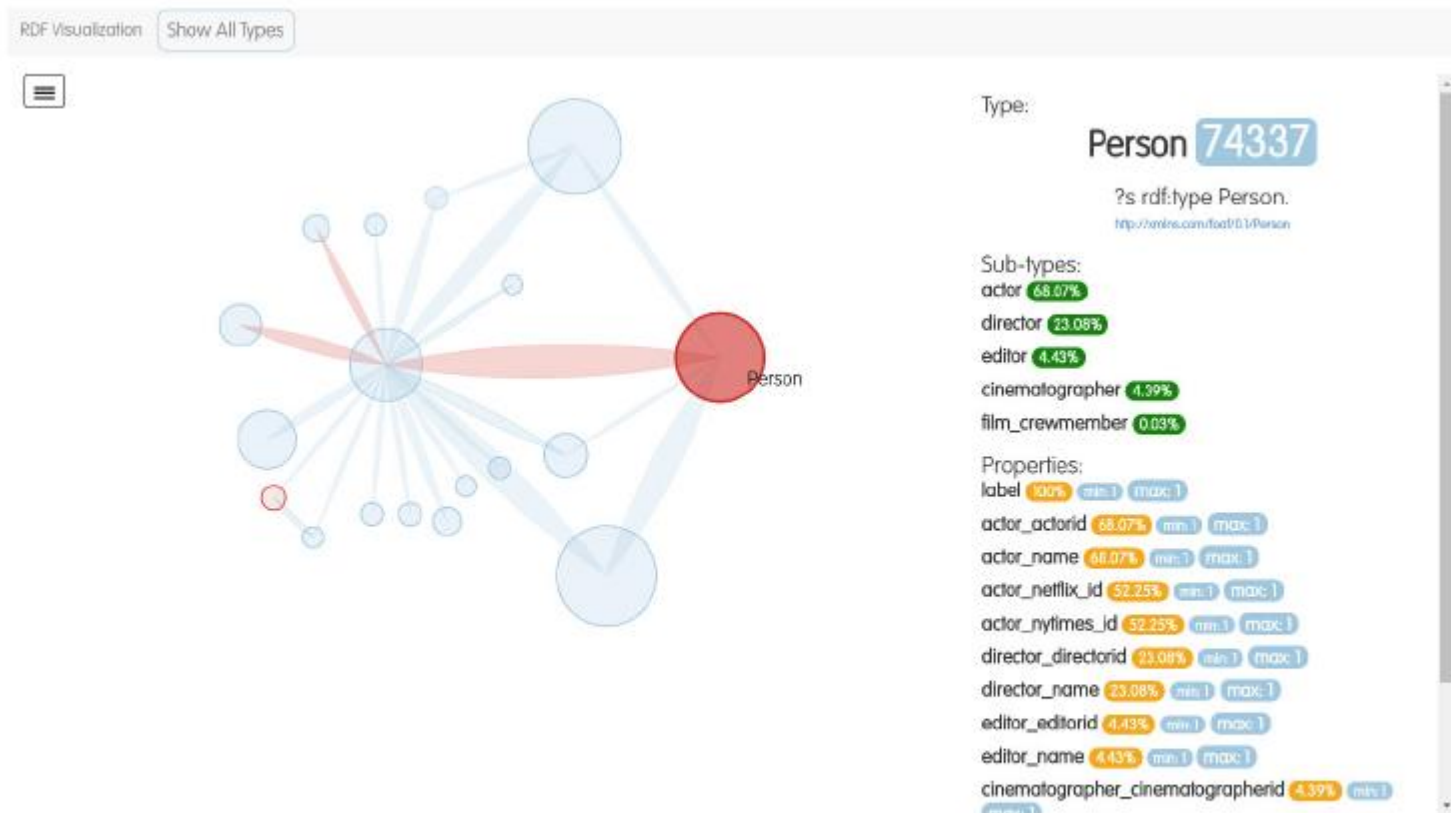
- 著者 : Fernando Florenzano, Denis Parra, Juan L. Reutter and Freddie Venegas
- SPARQLエンドポイントに対して, どんなタイプやプロパティを持っているか含有率とともに可視化表示できる
- クエリ生成も可能

Linked Data for Information Extraction (2日目開催)

■ Enhancing Open Data to Linked Open Data with ODMiner

- 著者 : Francesco Poggi, Andrea Giovanni Nuzzolese, and Gabriele Cigna
- CSV, XML, JSON等のオープンデータからLOD生成するODMiner
- 異なるオープンデータの分析を目的としたWebツール
- テーブルデータを読み込むと既存オントロジー語彙を基にカラムに意味を割り当てる
- RESTインターフェース

A Visual Aide for Understanding Endpoint Data



(a) Graph with Person node

参考文献

- [1] Andrea Giovanni Nuzzolese, Anna Lisa Gentile, Valentina Presutti, Aldo Gangemi.: Conference Linked Data: The ScholarlyData Project. In: Proc. ISWC2016, pp.150-158
- [2] Silvio Peroni, Giorgia Lodi, Luigi Asprino, Aldo Gangemi, Valentina Presutti.: FOOD: Food in Open Data. In: Proc. ISWC2016, pp.168-176
- [3] Amit Krishna Joshi, Pascal Hitzler, Guozhu Dong.: LinkGen: Multipurpose Linked Data Generator. In:Proc. ISWC2016, pp.113-121
- [4] Wei Hu, Haoxuan Li, Zequn Sun, Xinqi Qian, Lingkun Xue, Ermei Cao, Yuzhong Qu.: Clinga: Bringing Chinese Physical and Human Geography in Linked Open Data. In:Proc. ISWC2016, pp.104-112
- [5] Alberto Tonon, Victor Felder, Djellel Eddine Difallah, Philippe Cudré-Mauroux.: VoldemortKG: Mapping schema.org and Web Entities to Linked Open Data. In: Proc. ISWC2016, pp.220-228
- [6] Steven de Rooij, Wouter Beek, Peter Bloem, Frank van Harmelen, Stefan Schlobach.: Are Names Meaningful? Quantifying Social Meaning on the Semantic Web. In:Proc. ISWC2016, pp.184-199
- [7] Ilaria Tiddi, Mathieu d'Aquin, Enrico Motta.: Learning to Assess Linked Data Relationships Using Genetic Programming. In:Proc. ISWC2016, pp.581-597
- [8] Fernando Florenzano, Denis Parra, Juan L. Reutter and Freddie Venegas.: A Visual Aide for Understanding Endpoint Data. In: Proc. VOILA2016, pp.102-113
- [9] Francesco Poggi, Gabriele Cigna, Andrea Giovanni Nuzzolese.: Enhancing Open Data to Linked Open Data with ODMiner. In:Proc. LD4IE2016, pp.44-50
- *1 <http://data.semanticweb.org/>
- *2 http://data.semanticweb.org/ns/swc/swc_2009-05-09.html
- *3 http://ontologydesignpatterns.org/wiki/Main_Page
- *4 <https://github.com/anuzzolese/cLODg2>