

ISWC2018 参加報告

2018年10月18日

株式会社富士通研究所

小柳佑介

開催スケジュール

10/8	10/9	10/10	10/11	10/12
Tutorial & Workshop	Tutorial & Workshop	Opening Ceremony + Keynote	Keynote	Keynote
Morning Break				
Tutorial & Workshop	Tutorial & Workshop	各Session	各Session	各Session
Lunch				
Tutorial & Workshop	Tutorial & Workshop	各Session	各Session	各Session
Afternoon Break				
Tutorial & Workshop	Tutorial & Workshop	各Session	各Session	Closing and Awards
Dinner				
		Gala dinner	Posters & Demos Jam Session	

訪問スケジュール

10/8	10/9	10/10	10/11	10/12
Tutorial & Workshop	Tutorial & Workshop	Opening Ceremony + Keynote	Keynote	Keynote
		Morning Break		
Tutorial & Workshop	Tutorial & Workshop	各Session	各Session	各Session
		Lunch		
Tutorial & Workshop	Tutorial & Workshop	各Session	各Session	各Session
		Afternoon Break		
Tutorial & Workshop	Tutorial & Workshop	各Session	各Session	Closing and Awards
		Dinner		
		Gala dinner	Posters & Demos Jam Session	

4th Workshop on Semantic Deep Learning (SemDeep-4)

※Afternoon Break後から聴講

- Semantic WebのDeep Learningに関するワークショップ
 - ESWC2017の併設ワークショップから始まり4回目（1年に2回、今年2回目）
 - ※スライドと原稿: <http://www.dfki.de/semdeep-4/program.html>

発表

- *Combining Text Embedding and Knowledge Graph Embedding Techniques for Academic Search Engines*
 - IOS Pressの発表（論文検索のためにEmbeddingを活用）
- *Global and Local Evaluation of Link Prediction Tasks with Neural Embeddings*
 - *Link 予測のためにEmbeddingsを活用*
 - GitHub: <https://github.com/plumdeq/neuro-kglink.git>
- *Semantic Image-Based Profiling of Users' Interests with Neural Networks* [**Best Paper Award**]
 - ユーザの興味を画像からプロファイリング (デモ: <http://150.254.36.100:8080/>)
- *A First Experiment on Including Text Literals in KGloVe*
 - リテラル情報を考慮してEmbeddings

Tutorial: Challenges and Opportunities with Big Linked Data Visualization

- LOD可視化ツールのチュートリアル
 - スライド: <https://www.slideshare.net/polaura>
- ハンズオンセッションでは、Web上のLOD可視化ツール11種を紹介
 - Baloon <http://schlegel.github.io/balloon/balloon-synopsis.html>
 - Rhizomer <http://rhizomik.net/html/rhizomer/>
 - SynopsisViz <http://synopsviz.imis.athena-innovation.gr>
 - VizWizard <https://code.know-center.tugraz.at>
 - Aemoo <http://wit.istc.cnr.it/aemoo/>
 - GraphVizdb <http://83.212.97.26:8080/graphVizdb/>
 - LodLive <http://lodlive.it/>
 - RelFinder <http://www.visualdataweb.org/refinder/refinder.php>
 - Vowl (Queryvowl) <http://vowl.visualdataweb.org/queryvowl/queryvowl.html#>
 - LOG Linked Open Graph <https://log.disit.org/service/>
 - H-BOLD http://dbgroup.ing.unimo.it/hbold_bootstrap/
- 会場のネットワーク環境が貧弱で、現場での進行がかなり遅かった

Tutorial: Validating RDF data

- ShExとSHACLを用いたRDFのValidationのチュートリアル
 - <http://www.validatingrdf.com/tutorial/iswc2018/>
 - ESWC2016でも同様の内容でTutorialを実施していた
 - ※昨年、本を出している（HTML版あり）
 - <http://book.validatingrdf.com/>
- SHACLとShExとの違いがある程度明確化されていた（会場でもその質問が多かった）
- SHACLツールは、RecursiveなShapeを処理できず、RecursiveなShapeについて発表あり
（処理エンジンは未完成とのこと）

Oct 11, 14:00-15:20 (Session 3.4)

Blue Sky Ideas

DECENTRALIZING THE SEMANTIC WEB THROUGH INCENTIVIZED COLLABORATION

- セマンティックWebの分散管理についてのアイデア
- 発表者: Ruben Verborgh (Linked Data FragmentsやSolidプロジェクトに携わる人)
 - 今回、ワークショップDeSemWeb2018のorganizer、Comunicaの発表あり
- Tim Berners-Leeの昨年の三つのChallengesの二つ

Berners-Lee, T. (2017), "Three challenges for the Web, according to its inventor", World Wide Web Foundation, March, available at: <https://webfoundation.org/2017/03/web-turns-28-letter/>.

 - taking back control of our own data 1) We've lost control of our personal data
 - obtaining trusted information 3) Political advertising online needs transparency and understanding
- 各ノードでクエリを処理したり、部分的にdelegateしたりできる仕組み
- Blockchainで、TasksやIncentiveを記録

HOW MUCH IS A TRIPLE? (2位)

- Tripleの価値はいくらか、という発表
- 別セッションのBlue Sky Ideaで同じ発表者による「Make Embeddings Semantic Again!」という発表もあり
- 導出方法:
 - 全体での開発費用が分かっているものについては、単純にstatement数で割り算出
 - Freebaseは、Wikipediaの1sentence作成にかかった時間を算出して、最低賃金をかけて算出
- コストをかけるとエラー率も下がる(NELLは外れ値とのこと)

Oct 11, 14:00-15:20 (Session 2.4)

Ontologies & Linked Data

Resources track 4件 (BEST RESOURCE AWARD受賞1件)

※動画あり <https://www.youtube.com/watch?v=JZpTb8ropHY>

A FRAMEWORK TO BUILD GAMES WITH A PURPOSE FOR LINKED DATA REFINEMENT

- ゲームを活用して、Linked data/ナレッジグラフのカバレッジ向上と正確性向上
 - 機械学習やNLPの技術を使って解決するアプローチが提案されているが、ゴールドスタンダードな真実が必要となり、それは、人が手動で構築
- Linked data/ナレッジグラフのカバレッジ向上と正確性向上するためのゲームを作成するOSSフレームワークを紹介
 - GitHub: <https://github.com/STARS4ALL/gwap-enabler>
 - 実際に作られたゲームも三つReadMeからリンクされている
 - Link Ranking、Link Validation、Link Creation
 - チュートリアルも存在→数時間後には流用してゲームを作成可能？
 - GitHub: <https://github.com/STARS4ALL/gwap-enabler-tutorial>

DISTLODSTATS: DISTRIBUTED COMPUTATION OF RDF DATASET STATISTICS

- RDFデータセットのStatisticsを分散処理で算出するフレームワーク
 - AKSWのLODStatsのメンバーが著者に入っている
- DistLODStats
 - GitHub: <https://github.com/GezimSejdiu/Spark-RDF-Statistics>
 - SANSA(データフローエンジン)で動作 <https://github.com/SANSA-Stack>
 - LODStatsの7倍高速化
- 関連するポスター発表:
 - Apache Livyを使った、SANSA DistLODStatsのためのREST API
 - GitHub: <https://github.com/GezimSejdiu/STATisfy>

VOCALS: VOCABULARY & CATALOG OF LINKED STREAMS

※ポスターセッションでも発表があり、そちらではBestポスター受賞

- ストリームデータ固有のWebリソースを表現するための Vocabularyを提案
- 6つのrequirements
 - R1 enable the description of streams
 - R2 enable the description of streaming services
 - R3 enable historical stream processing/analysis and replay
 - R4 enable provenance tracking at any level (a) creation, (b) publication, (c) storage
 - R5 tame velocity for streaming data management
 - R6 tame variety for streaming data management
- 既存のLDS, SLD, LDNなどは、上記を全て満たすものは無いので提案（既存の合わせ技では、全て網羅されてはいる）

BROWSING LINKED DATA CATALOGS WITH LODATLAS

- LODAtlas: Linked Dataのカatalogポータル
 - <http://lodatlas.lri.fr/>
- メタデータと内容を合わせて視覚化し、興味のあるデータセットの探索を容易化
 - 内容: トリプル数、エンティティ数
 - メタデータ: 作成時間、更新時間

Oct 12, 11:00-12:00 (Session 1.4)

Industry

USE CASES OF THE INDUSTRIAL KNOWLEDGE GRAPH AT SIEMENS

- SIEMENSとmetaphactsによる発表
- KGを活用した五つのユースケースを紹介:
 - Building Digital Twin, Risk Management, Process Monitoring, Machine Service Operations, Factory Monitoring
 - ※内容は以下のPDFと同様
 - https://indico.cern.ch/event/669648/contributions/2838194/attachments/1581790/2499984/CERN_Open_Lab_Technical_Workshop_-_SIEMENS_AG_-_FISHKIN_-_11-01-2018.pdf
- Metaphactory Frameworkを活用
 - <https://www.metaphacts.com/product>

POPULATING THE FLE FINANCIAL KNOWLEDGE GRAPH

- FLE: Fujitsu Laboratories of Europe Ltd. (欧州富士通研究所)
- 複数のデータソースのデータから、金融分野での分析を手助け
 - 専門家の知見なく、新しいデータを既存のKGに結合
- 以下のプレスリリースに同様の内容の情報あり
 - <https://www.fujitsu.com/uk/about/resources/news/press-releases/2018/pr-fle20180612.html>

※発表動画: <https://www.youtube.com/watch?v=x0RA944RV4M>

KNOWLEDGE-BASED QUESTION ANSWERING FOR DIYERS

- Bosch社による発表
- DIYerのニーズに応答できるQAシステムを構築
(PDF原稿の図1参照)
 - Question Understanding→Grounding→Reasoning→Answer Generation
 - 質問の種類は、PDF原稿の表1参照
- 質問に必要な膨大な量のドメイン知識を基に知識ベースを構築

AMAZON NEPTUNE: GRAPH DATA MANAGEMENT IN THE CLOUD

- Amazon Neptuneについての発表
- 著者の何人かは、元Blazegraphの人(第一著者を含む)
 - Workshop: Scalable Semantic Web Knowledge Base Systems (SSWS)のInvited Talkでも話している
- ほとんどは以下の動画と同じスライド
 - Deep Dive on Amazon Neptune - AWS Online Tech Talks
<https://www.youtube.com/watch?v=rsAKj7sMbbQ>
 - AWS re:Invent 2017: NEW LAUNCH! Deep dive on Amazon Neptune (DAT318)
https://www.youtube.com/watch?v=6o1Ezf6NZ_E

Oct 12, 14:00-15:00 (Session 2.1)

Embeddings and Deep Learning

TOWARDS ENCODING TIME IN TEXT-BASED ENTITY EMBEDDINGS

- 時間の要素を考慮したEmbeddings方式
 - 筆者らが過去に提案したTEEによって各エンティティのベクトルを生成
 - 時間的実体(1943年など)に対し、関連するエンティティのベクトルを集約
- 時間を考慮した二つの類似度算出方式を提案
 - Time-Flattening: 時間要素を除外して類似判定
 - Time-Boosting: 時間要素を増やして類似判定
 - オバマ前大統領に類似した人物を算出すると、Time-Boostingで重みを増やしていくと最近の人物が上位に出てくるようになる

RULE LEARNING FROM KNOWLEDGE GRAPHS GUIDED BY EMBEDDING MODELS

- Embeddingモデルを使ったKGからの推論ルールの学習
 - GitHub <https://github.com/hovinhthinh/RuLES>
 - TextからのEmbedding Modelを生成し、ルールPruningのためのEvaluationに使用
- 検出されたルール例(原稿の表4)
`scriptwriter_of(X, Y) <- preceded_by(X, Z), scriptwriter_of(Z, Y), not tv_series(Z)`

A NOVEL ENSEMBLE METHOD FOR NAMED ENTITY RECOGNITION AND DISAMBIGUATION BASED ON NEURAL NETWORK

- NNに基づいたNERとNEDのENSEMBLE方式
 - ENNTR: Ensemble Neural Network for Type Recognition
 - ENND: Ensemble Neural Network for Disambiguation
- 各APIを統合
 - AlchemyAPI, DandelionAPI, DbSpotlight, TextRazor, Babelfy, MeaningCloud, ADEL, OpenCalais
- 四つのfeatureを使用 (Score, Entities, Types, Surface)

会場: Asilomar Conference Grounds



GALA Dinner/Jam session etc.

