

日本語版 DBpedia によるアノテーションを介した 要求と解決策の意味的関係の可視化

Visualization of the Semantic Relation between Requirements and Solutions using DBpedia Japanese and Annotation

張 确軒
Quexuan Zhang

早矢仕 晃章
Teruaki Hayashi

大澤 幸生
Yukio Ohsawa

東京大学大学院工学系研究科システム創成学専攻
Department of Systems Innovation, School of Engineering, the University of Tokyo

When we consider solutions for requirements, it is significant to make analysis on requirements and especially discover the implicit requirements behind them. However, this is usually not easy to perform, as a result that it is hard to find the radical solutions. In this paper, we propose a visualization method with the DBpedia Japanese as knowledge base to annotate the text of requirements and solutions, and show the semantic similarity among the linked concepts, which is designed to help us analyze requirements and make better solutions.

1. はじめに

新たなシステムや製品を設計するため、あるいは既存のシステムや製品を更新するため、要求分析は重要なプロセスである。要求分析活動には、まず要求を発見する必要がある。[Ohsawa 15]で提唱された Innovators Marketplace on Data Jackets というゲーム形式の創発支援技法により、参加者がお互いにコミュニケーションを取ってシナリオを創出しつつ、要求を提示したり解決策を提案したりすることで、従来にない要求や解決策が数多く出されてきた。しかし、その過程の中では、要求を理解する上で必要な情報がコミュニケーションから獲得できなかつたり、要求の中のある概念への認識が不十分であるという問題がある。その結果、要求分析が不完全になり、潜在要求を発見し損ねてしまう可能性がある。そして、次の解決策を提出する段階にも影響を与え、本質的な解決策に至らなくなる。そこで、本研究では、創発活動の中の要求分析と解決策の考案を支援することを目的とし、DBpedia[Lehmann 15]を知識ベースとして、要求と解決策の中の概念を顕在化し、それらの概念の間の意味的関係を可視化する手法を提案する。

2. 関連研究

DBpedia Spotlight は自動的に文の中の概念をアノテートし、非構造的な情報を Linked Open Data(LOD)と結び付けるアプリケーションである[Daiber 13]。しかし、日本語へのアノテーションには対応していない。

[Li 03]では、幾つかの意味的類似度の計算方法について考察を行われていた。Noia らは、LOD の `dcterms:subject` と `skos:broader` の二つのプロパティを用いて推薦システムの構築を試みた[Noia 12]。

3. 提案手法

まず DBpedia により、要求と解決策の文をそれぞれ概念のアノテーションを行う。要求の中の各概念に対して、リンクがある概念を取得し、解決策の中の概念と意味的可視化を行う。

3.1 概念名の獲得

日本語版 DBpedia のホームページから RDF データをダウンロードすることができる。概念名を獲得するため、`page-id` のファイルを用い、正規表現で概念名と ID を抽出し、TST 構造[Bentley 97]に保存する。生成した TST を用いて文の中の概念名を検出してアノテーションを行う。さらに、ID 情報も付与することができるため、後述の段階で保存する必要のあるデータのサイズを抑えることができる。

3.2 概念のデータ

SPARQL を利用し、DBpedia から色々な概念のデータを獲得することができる。

(1) リンク概念

インリンク概念というのは、ある概念を言及するリンク元の概念を意味する。反対に、アウトリンク概念とは、ある概念に言及されるリンク先の概念を指す。リンク概念を得るため、プロパティ `dbpedia-owl:wikiPageLink` を使う。主語と目的語を交換することにより、インリンク概念とアウトリンク概念を取得することができる。

(2) 意味階層

ある概念に対して、プロパティ `dcterms:subject` によってこの概念(リーフ概念)の分類概念を獲得することができる。分類概念に対して、`skos:broader` を使用すると、上層の分類概念を得ることができる。したがって、一つのリーフ概念から頂点の分類概念まで、各層の分類概念を取得でき、その経路も明らかにすることができる。また、分類概念の間に閉路が形成することがある。例えば、「知識」という分類概念の上層概念には、「学問」がある。「学問」の上層概念の中の一つは「知識」である。「知識」と「学問」はお互いに上層概念になっている。

3.3 意味的關係性

意味的關係性とは、二つの概念の意味階層による意味的類似度である。本稿では、二つの概念の最短経路の逆数を類似度と定義する。

3.4 可視化

概念に関するデータを備えた上で、要求と解決策の文に対してアノテートをし、インタラクティブな可視化を行う。

(1) 要求概念

要求の中のアノテートされた要求概念をクリックすると、図1のような可視化ビューが出てくる。円心にはクリックされた要求概念が位置する。右の緑色の半分はインリンク概念の集合であり、左の赤色の半分はアウトリンク概念である。各リンク概念と要求概念の距離(半径)はそれらの間の意味的類似度で決められている。類似度が高いほど半径が短い。ただし、図を見やすくするため、類似度を n 階層(図1では $n=5$)に分けてスケールングを行い、同じ層に属する概念を等間隔で配置した。表示されるリンク概念の数は、類似度の最大と最小の閾値を調節することによって変えることができる(図1では最大値は 0.32, 最小値は 0.13)。

リクエスト: 帰宅難民 になった時にどうやって帰ればよいか知りたい
 ソリューション: 東京 時の交通量を減らすの対策情報、主要地点の交通量、止まる可能性の高い交通手段との掛け合わせに、交通事故統計(起きる可能性と増える可能性がある)を加えて、確率的なリスト評価を行う。

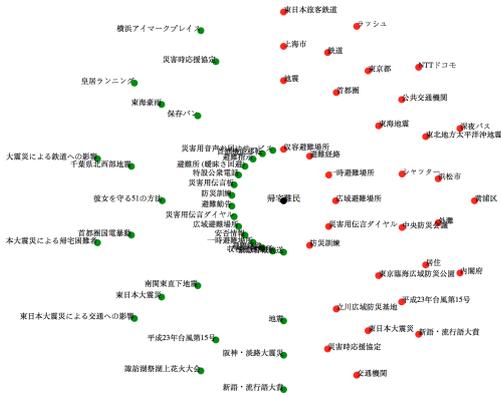


図1 要求概念「帰宅難民」をクリックした可視化結果

(2) 解決策概念

解決策の中でアノテートされた解決策概念にポインタを当てることで、図2に示されるように可視化ビューを変化させることができる。各リンク概念と指示された解決策概念の意味的類似度により、リンク概念の透明度が決められている。類似度が高いほど透明度が低い。他に調べたい解決策概念があれば、同じくポインタでその概念のアノテーションを指示すれば、各リンク概念の透明度が即時に解決策概念との類似度によって変化する。

リクエスト: 帰宅難民 になった時にどうやって帰ればよいか知りたい
 ソリューション: 東京 時の交通量を減らすの対策情報、主要地点の交通量、止まる可能性の高い交通手段との掛け合わせに、交通事故統計(起きる可能性と増える可能性がある)を加えて、確率的なリスト評価を行う。

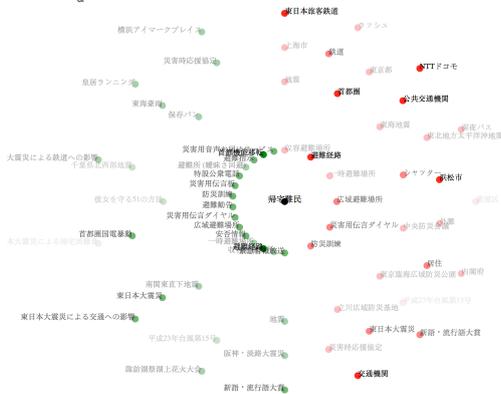


図2 解決策概念「統計」を指示した場合の可視化結果

4. 適用例

図1と図2に示された可視化結果を例として分析してみよう。まず「帰宅難民」という要求概念に対して、アウトリンク概念の中の「皇居ランニング」、「諏訪湖祭湖上花火大会」、「新語・流行語大賞」と、インリンク概念の中の「深夜バス」、「黄浦区」、「外灘」、「上海市」、「内閣府」などの概念は意外な繋がりをユーザーに提示できる。しかも、それらの概念は基本的に円の外縁に位置する。一方、円の内側のリンク概念は一見当たり前だと感じるが、見落とさないようにチェックリストとしても使える単語群が配置されている。図2では、解決策概念「統計」と意味的類似度が高い概念は、データとして用いられる可能性が高いではないかと推察される。要求分析と解決策の検討を行うとき、参加者はこの図を見つつコミュニケーションを行い、自身が保有する知識と連携し、潜在要求の発見と、より本質的な解決策の考案ができると考えられる。

5. 終わりに

日本語版 DBpedia により、要求と解決策の文に存在する概念をアノテートし、意味的な可視化を行った。本手法を利用すれば、要求の中に存在する概念を再認識することができ、解決策を提案する時に潜んでいた概念間の意味的関係性を直感的に認識できるだろうと考えられる。要求分析と解決策の考案をどれほど支援することができるかについては、今後の実験により検証を行う予定である。なお、今後 DBpedia の発展に伴い、より豊かな情報が得られることも期待している。

参考文献

[Ohsawa 15] Ohsawa, Y., Kido, H., Hayashi, T., Liu, C., & Komoda, K: Innovators Marketplace on Data Jackets, for Valuating, Sharing, and Synthesizing Data. *Knowledge-Based Information Systems in Practice*. Springer International Publishing, 83-97, 2015.

[Lehmann 15] Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P.N., Hellmann, S., Morsey, M., van Kleef, P., Auer, S., Bizer, C.: DBpedia – A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia. *Semantic Web Journal*, Vol. 6 No. 2, pp 167–195, 2015.

[Daiber 13] Daiber, J., Jakob, M., Hokamp, C., Mendes, P.N.: Improving Efficiency and Accuracy in Multilingual Entity Extraction. *Proceedings of the 9th International Conference on Semantic Systems (I-Semantics)*. Graz, Austria, 4–6, September 2013.

[Li 03] Li, Y., Bandar, Z.A., McLean, D.: "An approach for measuring semantic similarity between words using multiple information sources." *Knowledge and Data Engineering, IEEE Transactions on* 15.4, 871-882, 2003

[Noia 12] Di Noia, T., Mirizzi, R., Ostuni, V. C., Romito, D., & Zanker, M.: "Linked open data to support content-based recommender systems." *Proceedings of the 8th International Conference on Semantic Systems*. ACM, 2012.

[Bentley 97] Bentley, J. L., & Sedgewick, R.: "Fast algorithms for sorting and searching strings." *SODA*. Vol. 97. 1997.