人工知能学会研究会資料 SIG-SWO-046-03

# 動的オントロジーマッピング技術に基づく匿名化SPARQLクエリへの助言箇所の復元機構の実現

# Toward a Preliminary Approach for Deanonymizing Modifications and Comments using Dynamic Ontology Mappings based Query Anonymization

足立拓也 [1*]　福田直樹 [2]
Takuya Adachi[1]　Naoki Fukuta[2]

[1] 静岡大学大学院
[1] Graduate School of Integrated Science and Technology, Shizuoka University
[2] 静岡大学学術院情報学領域
[2] College of Informatics, Academic Institute, Shizuoka University

**Abstract:** 本研究では，匿名化に基づくSPARQLクエリ編集者と編集補助者との編集支援機構において，編集補助者が助言として改変した箇所をオントロジーマッピングを基づいて，匿名化される前のクエリへ復元する機構の実現について述べる．

## 1 はじめに

SPARQLのような構造化されたクエリ言語は表現力が高く，クエリ言語が持つ高度な機能を使いこなすための技術的なスキルとドメインスキーマの知識が求められる [Soylu 16]．そのため，様々な情報リソースに対して適切なクエリを記述することが難しい場合がある．初歩的なスキルや知識の不足を緩和し，Linked Open Data(LOD) から様々な情報を検索する試みとしては，キーワードベース，フォームベース，Faceted Searchなどの取り組みが行われている [Arenas 14][Ermilov 17]．また，ドメインスキーマの知識を緩和するための試みとしては，オントロジーマッピング [Noy 09] を用いたSPARQLクエリ記述の取り組みが行われている [Fujino 13][Makris 12]．

我々は，技術的なスキルが必要とされる複雑なクエリに着目し，複雑なSPARQLクエリの記述支援として，他者からの編集支援を受けるシナリオを想定する．SPARQLクエリ編集者が編集補助者からの編集支援を受ける機会があるとき，編集補助者はSPARQLクエリ編集者が公開している情報や編集支援を依頼したクエリを確認し，SPARQLクエリ編集者が取得したいクエリの対象を推測することができてしまう恐れがある．

*連絡先： 静岡大学
静岡県浜松市中区城北 3-5-1
E-mail: adachi.takuya.17@shizuoka.ac.jp

我々はSPARQLクエリ編集者と編集補助者との編集支援機構として，MCHA SPAIDAの実装を進めている [足立 18][Adachi 18b]．本機構では，SPARQLクエリを編集支援を依頼したクエリを編集補助を受ける目的を損なわない範囲で匿名化することを試みている．SPARQLクエリ編集者が匿名化SPARQLクエリを用いて編集補助者に質問したとき，編集補助者は匿名化SPARQLクエリに基づいて質問内容に対して助言を行う．編集補助者は，助言として質問内容に回答することや匿名化SPARQLクエリそのものに対して加筆・修正をすることができる．編集補助者からの助言を受け取った際，SPARQLクエリ編集者は本機構上でコメントや助言箇所が含まれた匿名化SPARQLクエリを閲覧する．

編集補助者が助言箇所を加えたクエリは匿名化したSPARQLクエリであり，SPARQLクエリ編集者が記述したクエリそのものには編集補助者の助言箇所は反映されていない．SPARQLクエリ編集者は編集補助者からの助言を閲覧しながら，SPARQLクエリ編集者が抱えていた疑問の解決を試みることになる．本稿では，助言箇所が含まれている匿名化SPARQLクエリから，匿名化SPARQLクエリを元のSPARQLクエリに復元し，助言箇所を元のSPARQLクエリに反映することを試みる．

## 2 MCHA SPAIDA

我々は SPARQL クエリ編集者と編集補助者との編集支援機構として，MCHA SPAIDA の実装を進めている [Adachi 18b]．MCHA SPAIDA は，SPARQL クエリ編集者が記述している SPARQL クエリに関する議論を行いたいとき，その議論内容と記述している SPARQL クエリを送信することで，編集補助者に編集支援を依頼し，議論する仕組みを検討している．記述している SPARQL クエリに関する議論を行う際，SPARQL クエリに公開したくない情報が含まれている場合において，SPARQL クエリを匿名化する仕組みを提案している [Adachi 18a]．

SPARQL クエリを匿名化する仕組みとしては，オントロジーマッピング [Noy 09] を用いて，SPARQL クエリ編集者が記述した SPARQL クエリから隠したい語彙を他の語彙に置き換えることを行っている．匿名化した SPARQL クエリに求められる条件としては，元の SPARQL クエリの対象を秘匿できることや，SPARQL クエリとして実行した場合において元の SPARQL クエリに似た傾向の結果を得られること，匿名化した SPARQL クエリの助言箇所をできるだけ元の SPARQL クエリに反映しやすいことであると考える．我々はこれらの条件を満たすため，Semantic Relatedness[Collins 75] と Multi-Layer Index[Liang 17] のような Graph Similarity の 2 つの尺度を用いて，SPARQL クエリ匿名化で使用するオントロジーマッピングを動的オントロジーオントロジーマッチング機構 [Adachi 17] を用いて生成している．

図 1 に MCHA SPAIDA での SPARQL クエリ匿名化の外観の例を示す．SPARQL クエリ編集者が編集補助者に議論を行いたいクエリを記述する．記述したクエリに SPARQL クエリ編集者が編集補助者に隠したい情報が含まれている場合，本機構は隠したい情報を匿名化するために動的オントロジーマッピングを生成する．動的オントロジーマッピングを生成した後，生成した動的オントロジーマッピングを基づいて議論を行いたいクエリを匿名化クエリに変換する．本機構では動的オントロジーマッピングの候補を複数生成し，SPARQL クエリ編集者の好みに合わせて匿名化クエリを選択できるようにしている．

## 3 MCHA メカニズム

我々は，オントロジーマッピングに着目し，オントロジーに含まれる語彙を用いて元のクエリから匿名化クエリに変換するための，オントロジーマッピングの生成手法について検討する．SPARQL クエリ匿名化におけるオントロジーマッピングの生成上の課題として，マッピング対象となるオントロジーの選定が挙げられる．SPARQL Endpoint Status[1]では，200 個以上のエンドポイントが利用できるとされており，また，Swoogle[2]では，10,000 個以上のオントロジーがあるとされている．

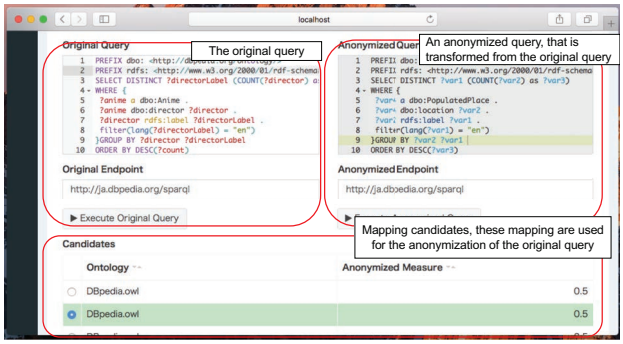


図 1: MCHA SPAIDA での SPARQL クエリ匿名化の例とその外観

SPARQL クエリでは，横断的な検索 (federated query) のような複数のオントロジーやスキーマを組み合わせて使用することができ，オントロジーによってはクラスやプロパティの構造が複雑に定義されているものがある．SPARQL クエリ中ではオントロジーの用語はごく一部しか使用されないため，オントロジーからサブグラフの取り出し方を検討すると，組み合わせ数の増加が懸念される．組み合わせ数が増加することで，計算コストや計算速度といった実装上の課題を解決する必要があると考えられる．匿名化の品質と実装上の課題との間にはトレードオフがある可能性があり，さらには，匿名化された SPARQL クエリの匿名化度合の好みはユーザごとに異なるため，ユーザに対して複数の匿名化 SPARQL クエリを示す必要があると考える．

### 3.1 SPARQL クエリ匿名化における尺度

我々は，SPARQL クエリの対象となるドメインの匿名化の尺度として，グラフ類似度と意味的類似度の 2 つを用いることを検討している．

グラフ類似度は Graph Similarity Search の分野で研究されており，ユーザが記述したグラフ構造のクエリが与えられたときに，適切なグラフを検索することができるようにすることが目的である [Liang 17]．グラフ間の類似度を計算する手法としては，グラフ編集距離 [Riesen 07]，Maximum Common Subgraphs[Shang 10]，Edge/Feature Misses[Yuan 15]，および Graph Alignment[Tian 07] が提案されている．MCHA メカニズムでは，グラフ編集距離に基づいて提案されている Multi-Layer Index(ML-Index) の利用を試みている．

[1]SPARQL Endpoint Status: http://sparqles.ai.wu.ac.at
[2]Swoogle: http://swoogle.umbc.edu/2006/

ここでは，Multi-Layer Index(ML-Index)[Liang 17] の概要を示す．文献 [Liang 17] では，グラフ $g$ を 4 つの要素 $(V_g, E_g, l_g, \sum)$ で構成している．ここで，$V_g$ はノードの集合，$E_g \subseteq V_g \times V_g$ はエッジの集合，$l_g : V_g \cup E_g \rightarrow \sum$ はラベルを特定する関数であり，$\sum$ はノードとエッジのラベルの集合である．ここで定義しているグラフ $g$ はグラフ編集を操作することができ，グラフ編集の操作としては，(1) ノードの追加，(2) エッジの追加，(3) ノードの削除，(4) エッジの削除，(5) ノードのラベルの変更，(6) エッジのラベルの変更といった 6 つ操作がある．2 つのグラフにおけるグラフ編集距離は，元となるグラフから対象となるグラフへとなるようにグラフを編集操作した際の最小の値で定義されている．つまり，2 つのグラフ $g$ と $g'$ が与えられたとき，グラフ $g$ をグラフ編集の操作を用いてグラフ $g'$ にする際の最小のグラフ編集操作総数である．しかしながら，グラフ編集距離の計算コストが NP-hard であると証明されており，ML-Index は効果的にハッシュを利用することによって計算コストを緩和するための手法である [Liang 17]．

意味的類似度（Semantic Relatedness）は 2 つの概念間の結びつきを表しており，意味的距離とはわずかに異なる観点から，2 つの用語を入れ替えて使われることがある [Collins 75]．Semantic Relatedness の具体的な手法として長年研究されており，主にコーパスベースの手法とグラフベースの手法の 2 つに分類することができる [Hulpuş 15]．コーパスベースの Semantic Relatedness は単語の分散表現や分散意味技術に基づいて計算される多次元ベクトルで表される [Mikolov 13][Pennington 14]．グラフベースの Semantic Relatedness は WordNet[3]や DBpedia[4]のようなグラフで構成されたナレッジベースに依存する．他のグラフベースの Semantic Relatedness の手法としてはオブジェクトプロパティを用いた手法 [Mazuel 08] やインスタンス間の関係性に基づいた手法 [Passant 10] といった外部知識を使用しない手法が提案されている．MCHA メカニズムでは，これらの手法の実装を進めており，オントロジーマッピングを生成する際にこれらの手法を使用できるようにしている．

## 3.2 MCHA SPAIDA の実装

我々は，MCHA SPAIDA 上にオントロジーマッピングに基づいた SPARQL クエリ匿名化手法の実装を進めている．図 2 に MCHA SPAIDA でのオントロジーマッピングにオントロジーマッピングに基づいた SPARQL クエリ匿名化手法の実際の使われ方を，具体的な質問

[3]Princeton University “About WordNet.” https://wordnet.princeton.edu/ Princeton University. 2010.

[4]DBpedia: https://wiki.dbpedia.org/

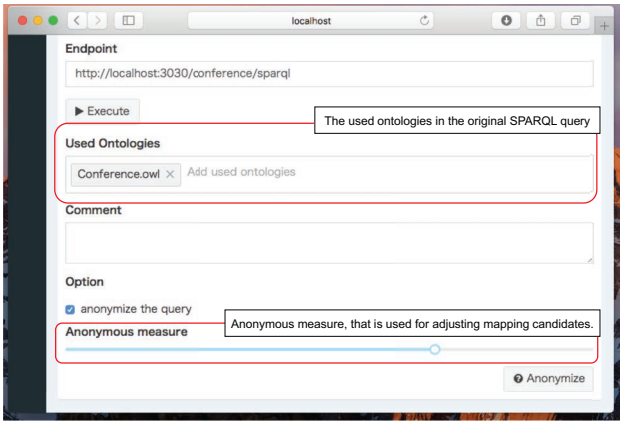


図 2: SPARQL クエリ匿名化における設定入力の例

内容と設定とともに示す．MCHA SPAIDA ではユーザインタフェース上のスライダーを使用することによってクエリ記述で用いたオントロジーと他のオントロジーとの意味的類似度を調整することができる．

図 3 に MCHA SPAIDA 上での SPARQL クエリ匿名化の流れを示す．SPARQL クエリ編集者が議論対象となる SPARQL クエリと SPARQL クエリを記述する際に用いたオントロジー，匿名化尺度などを入力すると，マッピング生成機構が SPARQL クエリを記述する際に用いたオントロジーとシステム上に格納されているオントロジーとのオントロジーマッピングを候補として複数個生成する．オントロジーマッピング候補を生成した後，SPARQL クエリ匿名化機構でオントロジーマッピング候補を用いることによって元のクエリから匿名化クエリへと変換する．我々のシステムでは，SPARQL クエリ編集者の好みに基づいて匿名化クエリを選択できるように，オントロジーマッピングの候補とともに匿名化クエリを確認することができる．

## 3.3 クエリ匿名化における課題

文献 [Adachi 18b] では，SPARQL クエリ編集者と編集補助者との SPARQL クエリ編集支援機構での SPARQL クエリ匿名化と復元におけるオントロジーマッピングを用いた手法を示した．そこでは，本編集支援機構における SPARQL クエリのプライバシーの保護を目的とし，SPARQL クエリ編集者がクエリ記述の際に使用していないオントロジーを用いることによって，元のクエリを意味的に類似している，または類似していないクエリにオントロジーマッピングを用いた変換手法を提案した．オントロジーマッピングに基づいた変換は，SPARQL クエリ編集者がクエリで検索したい事柄を匿名化するためにオントロジーマッピングを使用している．提案手法はグラフ類似度と意味的類似度の 2 つの

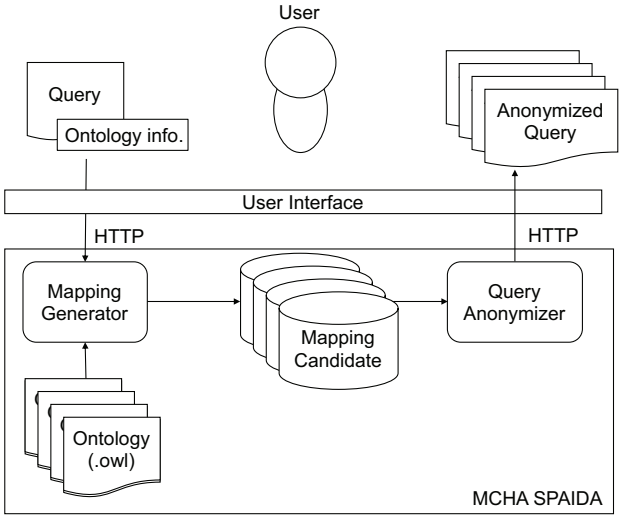


図 3: SPARQL クエリの匿名化

尺度を用いてオントロジーマッピングを複数候補として生成し，複数のオントロジーマッピングの候補から SPARQL クエリ編集者の好みに合わせて選択できるように匿名化 SPARQL クエリの候補を本編集支援機構上で確認することができる．

一方で，匿名化された SPARQL クエリに対する編集補助者の編集結果を適切に参照して匿名化される前の SPARQL クエリに反映し使用するには，匿名化される前と後の SPARQL クエリとそこで用いられるそれぞれのオントロジーに対する理解が必要となる場面があると考えられる．

## 4 助言箇所の匿名化復元機構

匿名化 SPARQL クエリから元の SPARQL クエリに復元する際，本機構で SPARQL クエリ匿名化の際に生成したオントロジーマッピングを使用する．このオントロジーマッピングは，SPARQL クエリ編集者が記述した SPARQL クエリにある語彙と，匿名先として選ばれた語彙とのマッピングである．SPARQL クエリ匿名化の際に生成したオントロジーマッピングを使用することによって，匿名化 SPARQL クエリから SPARQL クエリ編集者が記述したクエリへ復元することができる．

助言箇所として追記されたクエリ中の語彙を含むオントロジーマッピングが用意されていなかった場合，オントロジーマッピングを用いた復元機構では，助言箇所を SPARQL クエリ編集者が記述したクエリに反映することは難しいと考える．この対処として，本機構では助言箇所として追記されたクエリ中の語彙と SPARQL クエリ編集者が記述する際に使用したオントロジー中の語彙とのオントロジーマッピングを生成し，生成したオントロジーマッピングを用いて助言箇所を元の SPARQL クエリへ反映することを検討している．

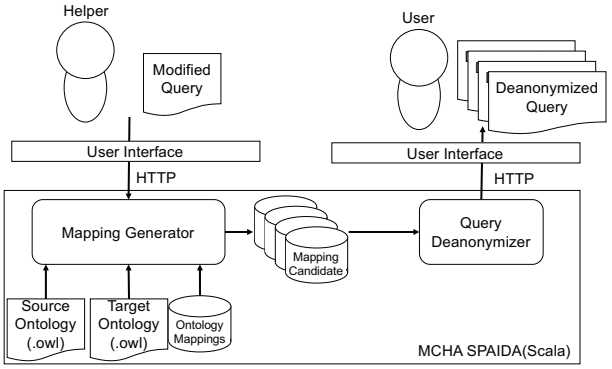


図 4: 匿名化 SPARQL クエリの復元

図 4 に匿名化 SPARQL クエリを復元する流れを示す．編集補助者から助言箇所が含む匿名化 SPARQL クエリが送信された際，本機構は送信されたクエリと，SPARQL クエリ編集者が SPARQL クエリを記述した際に用いたオントロジー，匿名化する際に対象となったオントロジー，匿名化する際に生成したオントロジーマッピングを用いて，助言箇所を反映するためのオントロジーマッピングを生成する．その後，生成したオントロジーマッピングを用いて助言箇所を含む匿名化 SPARQL クエリから助言箇所を含んだ元のクエリへと復元することを試みている．

図 5 に MCHA SPAIDA での匿名化 SPARQL クエリを元のクエリに復元する外観の例を示す．編集補助者が匿名化クエリに助言箇所として追記した場合，本機構ではその助言箇所を匿名化前のクエリに反映を試み，SPARQL 編集者にその結果を示す．匿名化クエリを匿名化前のクエリに復元するとき，本機構では SPARQL クエリを匿名化する際に生成したオントロジーマッピングに基づいて匿名化クエリを匿名化前のクエリへと変換する．編集補助者が助言箇所として新たに追加した語彙があった場合，本機構では新たにオントロジーマッピングを生成し，匿名化クエリを匿名化前のクエリへと変換する．このとき，オントロジーマッピングの候補を複数生成し，SPARQL クエリ編集者が望ましいと思われる匿名化クエリの復元をできるようにしている．

## 5 おわりに

本稿では，匿名化に基づく SPARQL クエリ編集者と編集補助者との編集支援機構において，編集補助者が助言として改変した箇所をオントロジーマッピングを基づいて，匿名化される前のクエリへ復元する機構について述べた．本機構では，SPARQL クエリ匿名化の

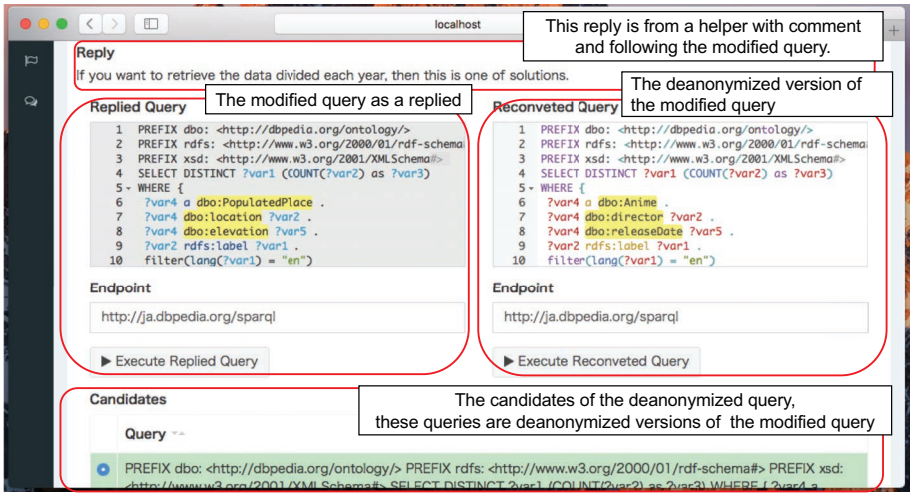


図 5: MCHA SPAIDA での匿名化クエリへの助言箇所の復元の例とその外観

際に生成したオントロジーマッピングを使用することによって，匿名化 SPARQL クエリから SPARQL クエリ編集者が記述したクエリへ復元する．

SPARQL クエリ匿名化手法の評価の指標となり得る観点としては，元の SPARQL クエリの対象を秘匿できていること，SPARQL クエリとして実行した場合において元の SPARQL クエリに似た傾向の結果が得られること，匿名化した SPARQL クエリの助言箇所をできるだけ元の SPARQL クエリに反映しやすいことの 3 つが挙げられる．

SPARQL クエリ編集者によって，匿名化 SPARQL クエリの好みが異なることも考えられるため，本機構では SPARQL クエリ編集者が好ましいと思われる匿名化 SPARQL クエリを選択できるようにしている．このため，本機構では複数の匿名化 SPARQL クエリの候補を提示する必要があり，匿名化 SPARQL クエリの候補数やその妥当性を計測することが，その評価の 1 つの手段として考えられる．また，匿名化 SPARQL クエリの候補が元のクエリに似た傾向の結果が得られるかを調べることができるようにする必要がある．これらの評価のため，グラフおよびクエリの生成器である gMark[Bagan 17] を用いて人工的にインスタンスデータとクエリを生成し，匿名化 SPARQL クエリの候補数と匿名化クエリが元のクエリに似た傾向の結果が得られるか，SPARQL クエリ匿名化の所要時間を計測していくことが，その 1 つのアプローチになると考える．

匿名化した SPARQL クエリの助言箇所をできるだけ元のクエリに反映しやすいことを評価するために，SPARQL クエリへの支援の具体的なユースケースを検討し，匿名化 SPARQL クエリで助言箇所を追記することで事前に想定した正当となるクエリが得られる度合いの計測やその特徴の分析が可能かを検証していくことも検討している．

## 謝辞

本研究の一部は，JST CREST JPMJCR15E1 の支援を受けたものである．

## 参考文献

[Adachi 17] Adachi, T. and Fukuta, N.: A Mapping-enhanced Linked Data Inspection and Querying Support System using Dynamic Ontology Matching, in *Proc. of 2nd International Workshop on Platforms and Applications for Social problem Solving and Collective Reasoning (PASSCR2017)*, pp. 1191–1194 (2017)

[Adachi 18a] Adachi, T. and Fukuta, N.: A Query Anonymization Approach using Ontology Mappings, in *Proc. of The Joint International Workshop on PAOS2018 and PASSCR2018* (2018), (to appear)

[Adachi 18b] Adachi, T. and Fukuta, N.: MCHA SPAIDA: A Cooperative Query Editor with Anonymous Helpers using Ontology Mappings, in *Proc. of The 13th International Workshop on Ontology Matching (OM2018)* (2018), (poster), (to appear)

[Arenas 14] Arenas, M., Grau, B. C., Kharlamov, E., Marciuška, Š., Zheleznyakov, D., and Jiménez-Ruiz, E.: SemFacet: Semantic Faceted Search over Yago, in *Proc. of the 23rd International Conference on World Wide Web (WWW2014)*, pp. 123–126 (2014)

[Bagan 17] Bagan, G., Bonifati, A., Ciucanu, R., Fletcher, G. H. L., Lemay, A., and Advokaat, N.: gMark: Schema-Driven Generation of Graphs and Queries, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 29, No. 4, pp. 856–869 (2017)

[Collins 75] Collins, A. M. and Loftus, E. F.: A Spreading-Activation Theory of Semantic Processing, in *Psychological Review*, Vol. 82, pp. 407–428 (1975)

[Ermilov 17] Ermilov, T., Moussallem, D., Usbeck, R., and Ngomo, A.-C. N.: GENESIS A Generic RDF Data Access Interface, in *Proc. of International Conference on Web Intelligence 2017 (WI2017)*, pp. 125–131 (2017)

[Fujino 13] Fujino, T. and Fukuta, N.: Utilizing Weighted Ontology Mappings on Federated SPARQL Querying, in *Proc. of the 3rd Joint International Semantic Technology Conference (JIST2013)* (2013)

[Hulpuş 15] Hulpuş, I., Prangnawarat, N., and Hayes, C.: Path-based Semantic Relatedness on Linked Data and its use to Word and Entity Disambiguation, in *Proc. of the 14th International Semantic Web Conference (ISWC2015)*, pp. 442–457 (2015)

[Liang 17] Liang, Y. and Zhao, P.: Similarity Search in Graph Databases: A Multi-layered Indexing Approach, in *Proc. of 2017 IEEE 33rd International Conference on Data Engineering (ICDE2017)*, pp. 783–794 (2017)

[Makris 12] Makris, K., Bikakis, N., Gioldasis, N., and Christodoulakis, S.: SPARQL-RW: Transparent Query Access over Mapped RDF Data Sources, in *Proc. of the 15th International Conference on Extending Database Technology (EDBT2012)*, pp. 610–613 (2012)

[Mazuel 08] Mazuel, L. and Sabouret, N.: Semantic Relatedness Measure Using Object Properties in an Ontology, in *Proc. of the 7th International Semantic Web Conference (ISWC2008)*, pp. 681–694 (2008)

[Mikolov 13] Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J.: Distributed Representations of Words and Phrases and their Compositionality, in *Proc. of International Conference on Neural Information Processing Systems (NIPS2013)*, pp. 3111–3119 (2013)

[Noy 09] Noy, N. F.: Ontology Mapping, in Staab, S. and Studer, R. eds., *Handbook on Ontologies*, pp. 573–590, Springer-Verlag Berlin Heidelberg (2009)

[Passant 10] Passant, A.: dbrec — Music Recommendations Using DBpedia, in *Proc. of the 9th International Semantic Web Conference (ISWC2010)*, pp. 209–224 (2010)

[Pennington 14] Pennington, J., Socher, R., and Manning, C. D.: GloVe: Global Vectors for Word Representation, in *Proc. of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP2014)*, pp. 1532–1543 (2014)

[Riesen 07] Riesen, K., Fankhauser, S., and Bunke, H.: Speeding Up Graph Edit Distance Computation with a Bipartite Heuristic., in *Proc. of International Workshop on Mining and Learning with Graph* (2007)

[Shang 10] Shang, H., Lin, X., Zhang, Y., Yu, J. X., and Wang, W.: Connected Substructure Similarity Search, in *Proc. of the 2010 ACM SIGMOD International Conference on Management of Data*, pp. 903–914 (2010)

[Soylu 16] Soylu, A., Giese, M., Jimenez-Ruiz, E., Vega-Gorgojo, G., and Horrocks, I.: Experiencing OptiqueVQS: a multi-paradigm and ontology-based visual query system for end users, *Universal Access in the Information Society*, Vol. 15, No. 1, pp. 129–152 (2016)

[Tian 07] Tian, Y., McEachin, R. C., Santos, C., States, D. J., and Patel, J. M.: SAGA: a subgraph matching tool for biological graphs, *Bioinformatics*, Vol. 23, No. 2, pp. 232–239 (2007)

[Yuan 15] Yuan, Y., Wang, G., Xu, J. Y., and Chen, L.: Efficient distributed subgraph similarity matching, *The VLDB Journal*, Vol. 24, No. 3, pp. 369–394 (2015)

[足立 18] 足立 拓也, 福田 直樹 : SPARQL クエリ編集者と編集補助者との匿名化マッチングおよび編集支援機構のオントロジーマッピング手法を用いた試作 , 第 44 回セマンティックウェブとオントロジー研究会 (SIG-SWO-044) (2018)