様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２８ 年　 ６ 月 １０ 日現在

機関番号：１０１０１
研究種目：基盤研究(B)（一般）
研究期間：2013～2015
課題番号：２５２８００３５

研究課題名（和文）Linked Open Dataを用いた固有名詞タグ付けと情報検索への応用

研究課題名（英文）Named entity recognition using Linked Open Data and its application to information retrieval

研究代表者

吉岡　真治（Yoshioka, Masaharu）

北海道大学・情報科学研究科・准教授

研究者番号：４０２９０８７９

交付決定額（研究期間全体）：（直接経費）　　8,500,000円

研究成果の概要（和文）：本研究では、Wikipedia上のメタデータとして整理したDBPediaの情報を利用した固有名詞とそのタイプ（人名・地名・組織名など）の抽出を行うとともに、そのタイプの情報を考慮した情報検索に有用なインデックスについて検討を行い、固有名詞のタイプを観点（facet）として扱い、複数の観点の存在を考慮したfacet-biasedトピックモデルを提案した。また、Wikipediaの情報の一貫性をDBPediaの情報を用いて分析するためのツールであるWC3を作成し、具体的に、Wikipedia中に存在するメタデータのエラーや、カテゴリ付与の一貫性がかける事例などを指摘できることを確認した。

研究成果の概要（英文）：In this research, we proposed a framework to extract named entity from the text with named entity type information by using DBPedia. We also proposed facet-biased topic model that take into account the type of named entity as facet. In addition, we implemented WC3 that analyzes consistency of metadata in articles that belongs to a particular Wikipedia category. We confirmed that WC3 can find out the articles that have inconsistent metadata and candidate articles for the category.

研究分野：情報検索

キーワード：情報検索　固有名抽出　Linked Open Data

１．研究開始当初の背景

情報検索において、固有名詞に関連する検索は非常に多く行われており、Google などのサーチエンジンを使うと、固有名詞から、その固有名詞に対応するオフィシャルサイトなどを簡単に見つけられる。一方で、オフィシャルサイト以外にある情報を網羅的に見つけたい場合や、固有名詞に関連する詳細な検索要求を持っている場合に、必ずしも、オフィシャルページが適切な情報源であるとは限らない。固有名詞を含む詳細な検索要求が必要となる場合としては、質問応答などの高度な言語処理を伴うシステムの前処理としての情報検索などが考えられる。この問題は、情報検索における主要な会議の一つである ACM-CIKM 2010 の基調講演においても指摘されている非常に重要な課題である。一方、これらの固有名詞の情報を用いた情報検索の方法論については、様々な方法が提案されているものの、まだ、決定的な解決策が提案されていない段階である。

また、固有名詞に対する知識の構築については、Web 上の百科事典である Wikipedia、地名情報のデータベースである GeoNames、語彙概念階層を含む辞書である WordNet に代表される Linked Open Data の整備にともない様々な形で行われている。代表的なものとしては、Wikipedia オントロジーと呼ばれるもので、YAGO2 や DBPedia (http://dbpedia.org/) などがあり、Wikipedia や Wikipedia, GeoNames, WordNet の複数の知識源を統合する手法も提案されているが、Wikipedia の情報の正確性などについての議論はあまり行われていない。

２．研究の目的

本研究では、一般の文書を対象に、固有名詞に関連する検索質問に対して、より適切な文書を検索するための枠組の提案を目標としている。この枠組の構築のために、Linked Open Data (LOD) の中心的な役割を果たす Wikipedia, DBPedia を利用した固有名詞辞書の作成を行うとともに、文書中から固有名詞のタイプに応じた情報を取り出し、そのタイプに応じた固有名詞の重要性を考慮した情報検索のためのインデックスについて検討を行い、具体的なシステムを構築する。また、LOD の中心となる Wikipedia の情報について、分析を行う。

３．研究の方法

本研究では、Linked Open Data の中心的な存在である DBPedia, 日本語版 DBPedia の情報を用いて、固有名詞のタイプに関する情報を付与した単語辞書を構築するとともに、既存の固有名詞抽出のツールと組み合わせることにより、文書からタイプ分類付きの固有名詞の抽出を行う。

このタイプ付きの固有名詞の情報を文章を表現する facet(観点)と考えて活用する方法として、LDA(Latent Dirichlet allocation)に基づくトピックモデルを拡張した facet-biased トピックモデルという新たなモデルを提案し、新聞記事のカテゴリ分類の問題において、提案した facet-biased トピックモデルの有用性を確認する。

また、Wikipedia のカテゴリの情報と、DBPedia により抽出されたメタデータの情報を比較することにより、Wikipedia に記述されているメタデータの信頼性やカテゴリ付与の網羅性を調査できるツールである WC3(WC-triple):Wikipedia Category Consistency Checker を作成する。本システムを用いることにより、DBPedia が主にデータの抽出に用いる Wikipedia のページ中の Infobox の情報と、Wikipedia のカテゴリの情報の一貫性を比較することにより、Infobox の情報が適切かどうか、Wikipedia のカテゴリが網羅的に付与されているかなどを調べることができ、Wikipedia のボランティア編集者を支援するツールとして提供している。

４．研究成果

（１）　facet-biased トピックモデル

トピックモデルとは、単語の生成確率により特徴づけられるトピック(例えば、スポーツに関係するトピックならば、スポーツに関連するトピックの単語の生成確率は高く、その他の単語の生成確率は低い)の組み合わせにより、各々の文書が生成されると考える文書の生成モデルであり、代表的な手法としては LDA がある。LDA では単語間に現れる単語間の共起情報を用い、全体の文書群中で特徴的に現れるような共起語群を中心としたトピックを生成し、各々の文書におけるトピックの混合比率を推定することにより、各々の文書の特徴を表すことが可能である。

facet-biased トピックモデルは、このトピックモデルを拡張し、トピックモデルが生成するトピックに、特定の facet(観点)とその関連語を生成するためのトピックという考え方を導入したトピックモデルである。このトピックモデルでは、facet 語は facet-biased トピックからのみ生成される（それ以外のトピックからの発生確率は 0、あるいは、0 に近い小さな値）と設定して LDA によるトピックの生成を行う。図１は、facet が２種類（赤色と青色）あり、単語の色がそれぞれ facet 語と facet-biased トピックに対応する。

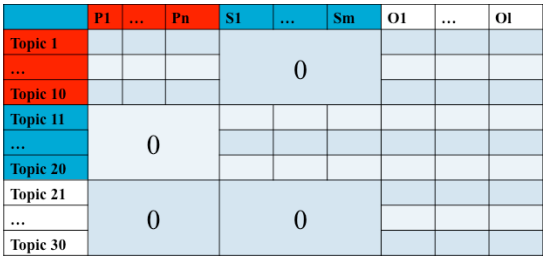

図１：facet-biased トピックモデルにおけるトピックとその単語の発生確率

この結果、全体の文書中に存在する各々の

facet 語が、その他の語の中に存在する特徴的な共起語を含む形でトピックに割り当てられ、各々の facet 語との共起性を考慮した分類が行われる。

本モデルの適用事例として、特許文書の対象(IC チップなどの対象語により表現)と観点(コスト、信頼性などの観点語により表現)という二つの facet を考えた特許マップの作成というタスクへの応用を行った。この結果得られた特許マップの事例を図２に示す。

		対象									
		ICカード	アンテナ	カード	情報システム	メモリ	ICチップとタグ	製造	システム	利用	認証装置
観点	信頼性	61	47	49	37	41	40	19	20	0	7
	耐熱性	47	44	23	17	0	28	19	0	0	0
	利便性	35	0	15	15	21	5	24	25	19	0
	生産性	40	23	38	0	1	15	2	7	0	4
	セキュリティ1	27	16	0	34	0	14	4	0	0	0
	セキュリティ2	26	0	0	2	41	1	0	0	3	1
	安全性1	4	0	6	19	0	1	12	0	0	1
	形状	24	0	1	0	1	1	6	0	0	1
	精度	0	10	0	1	0	0	0	0	0	4
	安全性2	0	0	0	3	0	0	0	0	0	0

図２：IC カードに関する特許マップ

図２中の対象・観点のラベルについては、得られた facet-biased トピックにおける発生確率の高い facet 語（例えば、対処の IC チップとタグというラベルは、IC チップ、IC モジュール、非接触式情報記憶媒体など）のリストを見てユーザが付与したものである。

本モデルを使うことにより、特許文書中の対象と観点の語を、対象語と観点語の共起情報を用いずに同時にクラスタリングするとともに、その関係を２次元のマップとして可視化することが可能となった。

（２） facet-biased トピックモデルと距離尺度学習を用いたニュース記事分類

本研究では、Yahoo! News (http://news.yahoo.co.jp)を対象データとして、既存のニュース記事のラベル情報に基づいて、新しいニュース記事のタイプ分類を行う問題に、facet-biased トピックモデルと、分類ごとに重視するトピックを考慮するための距離尺度学習を用いたシステムを構築した。

ニュースの記事分類では、固有名の存在が大きな役割を果たす。例えば、野球の始球式のニュース記事であっても、投げた人が芸能人であれば、芸能のトピックとして扱われ、アメリカ大統領が投げれば国際や政治の記事となる。また、地震に関する記事においても、日本で起きた地震は国内の記事になるが、海外で起きた地震は海外の記事となる。このような記事を分類するためには、一般的な語についての特徴を見るだけでなく、固有名に注目した特徴を考えることが有用であると考えた。

トピックモデルに代表されるような一般的な次元圧縮の手法では、このような固有名に注目した特徴を明示的に扱う方法は存在しない。そこで、本研究では、facet-biased トピックモデルの facet として、人名、地名、組織名という 3 つの facet を考え、これらの facet に関連するトピックを作成することにより、特徴的な固有名の共起性に基づくトピックが生成する。このような固有名に注目したトピックの混合比率として文書を表現することにより、固有名のタイプ(トピックとして表現される)を考慮した文書の特徴量が生成可能となる。

（１）で述べた特許マップの作成のための facet-biased トピックモデルでは、facet として 2 つの facet が利用されていたが、今回は固有名詞のタイプとして、人名、組織名、地名の 3 つの facet を利用する。

本研究では、Yahoo! News における記事分類（スポーツ、エンタメ、国際、国内、経済、地域、IT ・科学、ライフ）を対象に記事の分類を行う。各々の記事分類では、重視すべき固有名詞のタイプが異なることが考えらえるため、記事の分類時には、各カテゴリの分類に有用な距離尺度学習を LMNN (Large Margin Nearest Neighbor) により行うことにより、各分類に有用な距離尺度を作成し、実際の分類を行う。

本システムを用いて Yahoo! News で公開されているニュース記事を対象に、Yahoo! News が分類した記事分類を正解データとして、その正解データをどれくらい再現できるかという実験を行った。具体的には、2015 年 11 月 1 日から、2016 年 1 月 14 日にクロールしたデータを用い、4 週間分の記事を学習データ(平均記事数 62,648.6 件)として、次の 1 週間の記事分類を行うというテストデータ(平均記事数 14,642.4 件)を 5 組(学習データの先頭の日付が、11/1,11,21,12/1,11)作成し、各々の組について、2 つのトピックモデル（通常のトピックモデルもしくは facet-biased トピックモデル)、距離尺度学習の有無、2 つのトピック数(t=100,t=200)の組み合わせによる計 8 つの設定により実験を行った。

トピックモデルについては、scikit-learn の online batch 処理によるトピックモデルのライブラリを利用し、facet-biased トピックモデルについては、図１に示す単語の発生確率の制約を満すように、上記のライブラリを修正したものを利用した。また、トピックモデルに関連するパラメータについては、システムのデフォルト設定を利用した。

facet-biased トピックモデルで用いる人名・組織名・地名の推定については、Wikipedia のエントリーに基づく辞書で拡張をした MeCab と Cabocha を用いた固有名抽出システムを利用して、その推定を行った。また、その他の語としては、代名詞、接尾、数などを除く一般名詞を利用した。通常のトピックモデルを用いる場合においても、固有名も利用して文書ベクトルの作成を行った。

各々の文書に対応する文書ベクトルの作成時には、全体の文書群で 3 回以下しか現れない低頻度語を削除すると共に、文書中に存在する語に対応する重みとしては、TF-IDF を利用した。また、文書長を考慮した正規化を行うために、総和を 1 に正規化したベクトル

を利用して、トピックモデルの生成を行った。

また、facet-biased トピックモデルの生成時には、各々の facet に関係するトピックの数を設定する必要がある。本実験では、先の文書ベクトルの作成時に用いた各 facet 語の異なり語数に対応する形で、全体のトピックを分割することとした。例えば、人名:1000 語、地名：500 語、組織名：500 語、その他：3000 語で 100 トピックの場合は、人名：20 トピック,地名：10 トピック,組織名:10 トピック,その他:60 トピックと設定した。

トピックモデルにおける各文書の混合比率には、トピックに対応する語が存在しない場合においても、正則化のための非常に小さな定数比率を与えているが、本研究で考えているような文書分類の観点からは、この様な定数には意味がないと考え、これらの値を 0 として扱うこととした。また、先ほどのライブラリが出力する混合比率については、その大きさが正規化されていないため、先ほどの文書ベクトルの時と同じく、総和を 1 にする正規化を行った。

この結果得られたベクトルに対し、距離尺度学習を行わない場合は、そのまま kNN による分類を行い、距離尺度学習を行う場合には、LMNN のライブラリを用いて距離尺度学習を行いその結果を用いて kNN による分類を行う。ただし、LMNN の計算コストが非常に高いため、7000 記事(約 10%)のランダムサブサンプルを用いた計算を行った。また、kNN の k としては、5 を利用し、それ以外のパラメータはライブラリのデフォルト設定を利用した。

また、本データセットでは、ニュース記事の分類ごとに、記事数の大きな片寄りが存在する。そのため、記事数の大きなスポーツや芸能の分類性能は良く、記事数の少ないライフの分類性能が悪いという問題が発生した。この問題を解決するために、記事数の大きな分類から順番に記事分類を推定していく逐次的な分類を行うアルゴリズムを作成した。具体的には、次のような手順で分類を行う。

1. 記事数の多さを考慮して、分類の順序を「スポーツ→エンタメ→国際→国内→経済→地域→IT ・科学→ライフ」のように設定し、最初は、全ての学習データを対象学習記事、全てのテストデータを対象テスト記事、「スポーツ」を対象分類と設定する。
2. 対象学習記事を全て利用して、設定に応じたトピックモデルの学習、LMNN による学習(7,000 記事より多い場合には、7000 記事のサブサンプルを利用)を行い、kNN の分類器を作成する。この分類器を用いて、全ての対象テスト記事の分類を行い、対象分類と判定された記事の分類のみを確定する。ただし、分類対象が 2 つの場合には、全ての分類を確定し、最終結果とする。
3. 対象学習記事から、対象分類に属する記事を取り除いたものを新たな対象学習記事とし、分類の確定しなかった記事群を次の段階の対象テスト記事とする。分類の順序に応じて、次の分類(「スポーツ」の次は「エンタメ」)を対象分類と設定し、2.の手順に戻る。

この様な手順を繰り返すことにより、記事数の少ない分類に対しても、その分類を確定する際には、分類に関連するトピックが十分存在する中で分類が行われるだけでなく、各々の分類に適切な距離尺度学習が行われることが期待される。

トピック数 N を 100 と 200 に設定した場合、LMNN を利用した場合と利用しない場合、facet-biased トピックモデルを使った場合と従来型のトピックモデルを使った場合の８通りの組み合わせについて５組の実験データについて実験を行った平均を、表１に示す。

表１：ニュース記事の分類精度

N	**facet-biased**		**従来型**	
	LMNN	-	LMNN	-
100	**0.778**	0.772	0.734	0.734
200	**0.778**	0.770	0.735	0.734

この結果から、facet-biased トピックモデルを用いたほうが、従来型のトピックモデルを用いる場合よりも、統計的に有意な差があることをウィルコクソンの符合つき順位和検定により確認した($p<0.01$)。

（３） WC3(Wikipedia Category Consistency Checker)

Web 上の百科事典である Wikipedia には、多種多様な事象に関するページが存在している。このページの多くには、Infobox と呼ばれるページの内容のタイプに特有の構造化情報を表示する部分や、分類を表すためのカテゴリの情報などが付与されている。DBPedia は、各々のページからこの構造化された情報を抜出し、大規模な事象に関する構造化情報のデータベースを構築している。また、この DBPedia は、Linked Open Data の中心として、様々なデータと関連づけられて利用されている。

この DBPedia の情報の品質は、Wikipedia の記述に依存するが、その記述の品質については、編集者に依存する。この Wikipedia の記述に関する信頼性については、その記述内容についての分析や、ページの編集に携わった人々に関する属性を用いた分析などが行われている。また、DBPedia の情報については、他の Linked Open Data などと比較した分析なども行われている。しかし、DBPedia の情報を用いて、Wikipedia の情報を分析し、Wikipedia の品質向上につなげようという研究はほとんど行われていない。本研究では、DBPedia により抽出された構造化情報を用いて、Wikipedia のカテゴリ構造の一貫性を分析する方法を提案する。

Wikipedia のカテゴリ情報は、主に、ページの閲覧性の向上を目的として、類似した内容を含むページを、その内容を表す名前を持

つカテゴリことを目的として付与されている。このカテゴリは、ある種の包含関係を考慮した親子関係により、束上の階層関係を構成している。このカテゴリには、「日本」、「ポール・マッカートニー」といったトピックを表すようなカテゴリ、「作家」、「歌」などのクラスを表すようなカテゴリ、「日本の作家」などのトピックとクラスの組み合わせによりあらわされるカテゴリが存在する。特に、クラスに関連するカテゴリ(例えば、「作家」)については、一つのカテゴリに、あまりに多くのページが属する場合に、このカテゴリを分割した形であるトピックとクラスのカテゴリ(例えば、「日本の作家」「英国の作家」)を作ることが推奨されており、多くのページを持つようなクラスのカテゴリに関しては、様々なトピックとクラスの組み合わせのカテゴリが作られている。

本研究では、この考えに基づき、トピックとクラスであらわされるようなカテゴリについて、DBPedia のデータを用いることにより、その一貫性を検討する方法を提案する。具体的には、カテゴリに属するページが共通して持つ属性を用いて、カテゴリに属するページをできる限り過不足なく検索できる SPARQL のクエリの作成を行う。このような、SPARQL のクエリにより、カテゴリに属するページを過不足なく検索できる場合には、一貫した構造化情報が各ページに存在することが確認できる。

一方、完全に過不足ないクエリが作れない場合には、クエリの妥当性について検討するとともに、クエリを満たすがカテゴリに対応しないページや、カテゴリに属するがクエリを満たさないページについての分析が必要である。前者のページについては、本来、カテゴリに属すると判断してよいページに、適切なカテゴリが付与されていない可能性があり、後者のページについては、DBPedia が抽出可能な適切な構造化情報が存在しない可能性がある。これらの情報は、カテゴリ付与の一貫性を検証するための有用な情報となると考えている。

本研究では、カテゴリに属するページが持つ属性情報に基づいて、適切な SPARQL クエリを以下の手順により作成する。

1. カテゴリを入力とし、そのカテゴリに属するページ集合から他のページへのリダイレクトとなっているページを除いた集合 *Pc* を抽出する。
2. 抽出したページから 50 件を上限としてページを抽出して制約を表す属性候補の作成に利用する。また、トピックやクラスを表すキーワードを用いた FILTER 構文による制約も作成する。この構文で利用するキーワードの作成の際には、兄弟カテゴリとの文字列比較を行い、共通部分を除去することで、トピックやクラスを表すキーワードを作成する(例:「山下達郎の楽曲」について、兄弟カテゴリ「槇原敬之の楽曲」などを用いると、「山下達郎」を抽出し、「山下達郎のアルバム」を用いると、「楽曲」を抽出する)。
3. さらに、クラス属性を表すと考えられる http://www.w3.org/1999/02/22-rdf-syntax-ns¥#type と http://dbpedia.org/shortDescription を持つものについてはクラスの候補となる属性とする。
4. ２で作成したトピックに関する属性に関する制約と３で作成したクラスに関する属性の制約を組み合わせて候補となる SPARQL クエリを作成し、精度、再現率、F 値（精度と再現率の調和平均）を計算し、F 値の最も高いものをクエリの候補とする。

本システムを用いた解析事例を以下に示す。

例えば、「1973 births」というカテゴリを入力とした場合には、以下のような SPARQL クエリが作成された。

```
SELECT DISTINCT ?s
 WHERE {
 ?s rdf:type http://xmlns.com/foaf/0.1/Person .
 ?s dbo:birthDate ?o1 .
 FILTER regex (?o1, “1973”)
 }
 MINUS { ?s dbo:wikiPageRedirects ?o . }}
```

このクエリにより、カテゴリに属する 9,904 ページ中の 9,647 ページを検索することができ、精度、再現率ともに 0.97 となった。クエリを満たすがカテゴリに属さない 298 ページの中には、カテゴリ情報が付加されていない場合や、Infobox などに誤りがある場合などが存在した。また、カテゴリに属するがクエリを満たさない 257 ページについては、その多くが Infobox が存在しないために、DBPedia の情報が不十分である場合が多いことが確認された。

本システムは、英語版の DBPedia に基づく英語版と日本語 DBPedia に基づく日本語版を作成し、公開中である。

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計　５　件）

1. Masaharu Yoshioka and Noriko Kando: Comparative Analysis of GDELT Data Using the News Site Contrast System. In Proceedings of the First International Workshop on Recent Trends in News Information Retrieval co-located with 38th European Conference on Information Retrieval (ECIR 2016), pp. 63-65, 2016. (査読有)
2. Masaharu Yoshioka and Rhett Loban: WC3: Wikipedia Category Consistency Checker Based on DBPedia. In Proceedings of the 11th Intl. Conf. on Signal-Image Technology & Internet-Based Systems, 712--718,

2015. （査読有）

3. Masaharu Yoshioka: Analysis of Japanese Wikipedia Category for Constructing Wikipedia Ontology and Semantic Similarity Measure. In Information Retrieval Technology 10th Asia Information Retrieval Symposium, AIRS 2014, Kuching, Malaysia, December 3-5, 2014 Proceedings, pp. 588-598, LNCS8870, 2014. （査読有）
4. Xiao Hu and Noriko Kando: Evaluation of Music Search in Casual-Leisure Situations, In Proceedings of the 5th International Symposium on Information Interaction in Context (IIiX 2014) Workshop on Search for Fun (S4F 2014), 2014. （査読有）
5. Masaharu Yoshioka, and Takahiko Fujiwara: Construction of a Japanese Gazetteers for Japanese Local Toponym Disambiguation. In Proceedings of the 7th Workshp on Geographic Information Retrieval, pp. 57-63, 2013. （査読有）

〔学会発表〕（計 11 件）

1. 小野寺大輝，黄楽，吉岡真治：facet-biased トピックモデルと距離尺度学習を用いたニュース記事の分類．2016 年度人工知能学会全国大会，4B4-5，北九州国際会議場，北九州市，2016 年 6 月 6-9 日．
2. 小野寺大輝，吉岡真治：対象-観点を考慮した facet-biased トピックモデルと特許マップへの応用．言語処理学会第 22 回年次大会発表論文集，P13-5，東北大学，仙台市，2016 年 3 月 8-10 日．
3. 吉岡真治：日本語版 WC3(Wikipedia Category Consistency Checker) ― 日本語版 Wikipedia のカテゴリに所属するページのメタデータの一貫性の分析―．人工知能学会第 37 回セマンティックウェブとオントロジー研究会，SIG-SWO-037-04，慶応大学日吉キャンパス，横浜市，2015 年 11 月 13 日．
4. 吉岡真治，神門 典子：複数国の新聞からの多観点比較による分析～GDELT データを用いた分析～．人工知能学会合同研究会優秀賞記念講演（招待講演），慶応大学日吉キャンパス，横浜市，2015 年 11 月 13 日．（学会発表 8 の受賞記念講演）
5. Masaharu Yoshioka and Masahiko Itoh: Interactive Multidimensional Data Visualization based on Multi Dimensional Scaling with Distance Metric Learning, The 10th International Workshop on Information Search, Integration, and Personalization (ISIP 2015), North Dakota, USA, October 1-2, 2015.
6. Thaer M. Dieb, and Masaharu Yoshioka: Comparison of different strategies for utilizing two CHEMDNER corpora, In Proceedings of the Fifth BioCreative Challenge Evaluation Workshop, pp. 110-115, Sevilla, Spain, 2015 年 9 月 9-11 日
7. 吉岡真治，Rhett Loban：DBPedia の情報に基づく Wikipedia のカテゴリ情報の一貫性の分析．2015 年度人工知能学会全国大会(第 29 回)論文集，1G4-2，はこだて未来大学，函館市，2015 年 5 月 30 日-6 月 2 日．
8. 吉岡真治，神門典子：複数国の新聞からの多観点比較による分析～GDELT データを用いた分析～．インタラクティブ情報アクセスと可視化マイニング研究会 第 8 回研究会研究発表予稿集，慶応大学日吉キャンパス，横浜市，2014 年 11 月 21 日．（人工知能学会研究会優秀賞）
9. 吉岡真治：Wikipedia のカテゴリー階層関係の分類を用いた日本語 Wikipedia オントロジーの分析．2014 年度人工知能学会全国大会（第 28 回）論文集，CD-ROM 2J3-4，姫銀ホール，松山市，2014 年 5 月 12-15 日．
10. 岸桂太，吉岡真治：特長表現に注目した対象-観点型特許マップの自動生成．情報処理学会情報基礎とアクセス技術研究会，2014-IFAT-114-9，産総研臨海副都心センター，東京，2014 年 3 月 29 日．
11. 岸桂太，吉岡真治：特長表現に注目した特許マップの自動生成．情報処理学会情報基礎とアクセス技術研究会，2013-IFAT-111-10，北海道大学，札幌市，2013 年 7 月 22-23 日．

〔図書〕（計 0 件）

〔産業財産権〕

○出願状況（計 0 件）

○取得状況（計 0 件）

〔その他〕

ホームページ等

WC3(WC-triple):Wikipedia Category Consistency Checker
http://wnews.ist.hokudai.ac.jp/wc3

６．研究組織

(1)研究代表者

吉岡 真治（YOSHIOKA, Masaharu）

北海道大学大学院・情報科学研究科・准教授

研究者番号：40290879

(2)研究分担者

神門 典子（KANDO, Noriko）

国立情報学研究所・情報社会相関研究系・教授

研究者番号：80270445

(3)連携研究者