情報処理学会論文誌　Vol. 52　No. 12　3423–3434 (Dec. 2011)

質問応答のための情報検索への応用を目的とした確率型検索モデルとブーリアン検索モデルの組合せ

吉　岡　真　治[†1]

効率的な文書データベースに基づいた質問応答システムを構築するためには，情報検索システムの性能の向上が不可欠である．ここで，特定の人や組織に関する質問応答を目的とする場合を考えると，対象となる文書に，これらの人や組織に関する情報が入っていることは不可欠であり，ブーリアン検索モデルに基づいた検索式を利用することが望ましい場合がある．本研究では，人名や組織名に多く存在すると考えられる表記バリエーションのリストと動詞の同義語を考慮することによって，質問応答を考慮した適切なブーリアン検索式を構築するとともに，確率型検索モデルとブーリアン検索モデルを組み合わせた情報検索システム ABRIR（Appropriate Boolean query Reformulation for Information Retrieval）を利用したシステムを提案する．また，本システムを地理情報などに関する質問応答のタスクである NTCIR-8 GeoTime タスクに適用し，その有効性を検証した．

On a Combination of Probabilistic and Boolean IR Models for Question Answering

Masaharu Yoshioka[†1]

In order to make a good question answering (QA) system based on text database, it is better to have a good information retrieval (IR) system. To make a good IR system for QA about particular named entities, it is better to use Boolean IR model by using appropriate Boolean query. In this paper, we propose to use ABRIR (Appropriate Boolean query Reformulation for Information Retrieval) that constructs appropriate Boolean query by using list of synonyms and variation of Japanese katakana description. Evaluation results shows that ABRIR works effectively for the task of IR for QA (NTCIR-8 Geotime).

1. はじめに

質問応答（Question Answering）とは，自然言語で与えられた質問（たとえば，「富士山の高さは？」とか「坂本竜馬とはどんな人ですか？」）に対する答えを，文書データベースから探しだして答えるという課題である．一般に，質問応答のシステムでは，関連する文書を見つけ出すという情報検索の段階と，文書から答えを見つけ出すという情報抽出の段階の 2 つの段階に分けて構成されることが多い．また，一般的に，後者の情報抽出の段階では，構文解析やパターンマッチといった複雑な言語処理を要求することが多いため，初期段階での情報検索システムの検索性能を向上させることが，システム全体の性能向上に貢献すると考えられている．このような意識のもとに，質問応答のための情報検索（Information Retrieval for Question Answering）[1)–3)] という課題が設定され，様々なシステムが提案されている．

提案されているシステムについて調べたところ，ほとんどのシステムにおいて，確率型検索モデル・言語モデルを利用した検索モデル・ベクトル空間検索モデルといった，検索質問との部分照合を行う検索モデルを利用していた．

これに対し，ブーリアン検索式を構築するアプローチも存在する[4)] が，初期検索式から構築されるブーリアン検索式の検索結果が 0 件だった場合に，ブーリアン検索式を同義語によって拡張するという方法である．しかし，最終的な検索式にブーリアン検索式を用いているため，この手法を適用して，検索を行う場合には，再現率に問題があった．

これに対し，部分照合の検索モデルである確率型検索モデルにブーリアン検索モデルの考え方を組み合わせた情報検索システム ABRIR（Appropriate Booleanquery Reformulation for Information Retrieval）[5)] が提案されている．本システムは，初期検索式と，初期検索式による検索結果の上位で作成する疑似適合文書を比較することによって，適切なブーリアン検索式を構築するとともに，ブーリアン検索式とマッチしない文書のスコアにペナルティを与えるという形で，2 つの検索モデルの組合せを実現している．本システムを組み合わせることにより，再現率の問題が解消できると考えた．

また，本研究では，質問応答（特に，特定の固有名詞に関する質問応答）のための情報検索においては，特定の固有名詞の情報を含まない文書は，たとえ，周りのキーワードが似て

†1 北海道大学
Hokkaido University

© 2011 Information Processing Society of Japan

いても，役に立たない文書と考えられるという仮説に基づき，キーワードのタイプ分類を行ったうえでブーリアン検索の考え方を応用することがシステムの性能向上に寄与すると考えた．

本論文では，Web 文書検索のために作成された ABRIR について簡単に紹介した後に，今回の質問応答のための情報検索に応用するための手法を提案する．また，本システムを，地理情報に関する質問応答のための情報検索タスクである NTCIR-8 における GeoTime タスクのデータに適用したところ，単純な確率型検索モデルを用いたシステムより，有意に性能が向上することが確認された．

2. ABRIR（Appropriate Boolean query Reformulation for Information Retrieval）

2.1 検索モデルの特徴を生かした検索要求の表現

情報検索のモデルとして，代表的な手法に，本手法が用いる確率型検索モデルに加え，ベクトル空間モデル，ブーリアン検索モデルなどがある．確率型検索モデルおよびベクトル空間検索モデルとブーリアン検索モデルの一番の違いは，ユーザの検索式に対する考え方にある．

確率型検索モデルおよびベクトル空間モデルはユーザが与えた検索式に含まれる語集合と文書中に含まれる語集合の重なりに注目したモデルとなっている．そのため，場合によっては，ユーザが欲しいと思って書いた検索語が存在しなくても上位の適合文書として提示されることがある．これは，検索式との部分照合（Partial Match）を行うという考え方と考えることができる．これに対し，ブーリアン検索モデルでは検索式との完全照合（Exact Match）を行うことを考えているため，ユーザが欲しいと思って書いた検索語が存在していない場合は，適合文書から除外される．

よって，確率型検索モデルおよびベクトル空間モデルは，ユーザが検索要求を検索語としてうまく表現できないということを前提にしたモデルであり，ブーリアン検索モデルはユーザが検索要求をうまく表現できるということを前提としたモデルと考えることができる．

しかし，一般に，ブーリアン検索式の作成は容易ではない．たとえば，多くの検索語を and の条件を加えることにより，検索式を満たす文書が極端に少なくなったり，高頻度語を用いた or 条件を設定することにより，ブーリアン検索式による絞り込みの効果が少なくなり検索式を満たす文書が多くなりすぎたりすることという問題が起こる[6]．よって，ブーリアン検索モデルを利用するためには，ユーザがブーリアン検索式を作成する際の支援が不可欠であると考える．

2.2 適合文書を利用したブーリアン検索式の修正

ABRIR[5] では，すべての適合文書には，検索者の要求を表現する検索語が必ず含まれているという仮定に基づき，適合文書に応じた形でブーリアン検索式を修正する方法を提案する．この修正手順は以下のとおりである．ただし，現在のシステムでは，演算子として and と or のみを取り扱うこととし，not については考慮していない．

(1) 初期検索の実行
初期のブーリアン検索式の入力に基づき，初期検索を行う．ただし，ユーザが最初に入力した初期ブーリアン検索式は不十分であることが多いため，初期検索では，それ以外の and や or の演算子については，無視をして，確率型検索モデルに基づいた検索を行う．

(2) （疑似）適合文書の選択
初期検索の結果から，適合文書を選択する．ユーザとのインタラクションを仮定しない場合には，確率型検索モデルに基づいた初期検索の上位 5 件を疑似適合文書として選択する．ユーザとのインタラクションを仮定する場合には，ユーザによる適合文書の選択を求める．また，ユーザとのインタラクションを仮定する場合には，追加で，ブーリアン検索式を厳密に適用してフィルタリングを行った結果の提示と，適合文書の選択を行ってもよい．

(3) ブーリアン検索式の修正（**図 1**）
得られた（疑似）適合文書と初期ブーリアン検索式を比較し，次の手順により，ブーリアン検索式の修正を行う．

(a) 検索語の網羅性のチェック
検索語として用いた語がすべての適合文書に網羅的に含まれているかどうかを調べる．網羅的に含まれている場合には，ブーリアン検索式を作成するために利用できる検索語とする（図 1 の “A” と “C”）．

(b) or 要素の追加
基本的には，(a) の手順で求めた検索語を and 演算子でつないだブーリアン検索式を作成するが，初期ブーリアン検索式において，先の手順ですべての適合文書に網羅的に含まれていた検索語を，or 演算子と組み合わされて利用されていた場合（図 1 の “C” に対する “C or D”）には，対応する部分を検索語と置き換える．

© 2011 Information Processing Society of Japan

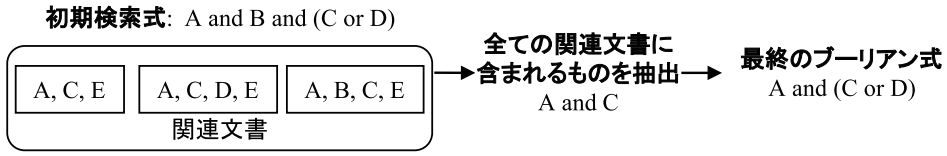

図 1　ブーリアン検索式の修正
Fig. 1　Boolean query construction.

2.2.1　情報検索におけるブーリアン検索モデルと確率型検索モデルの統合

ブーリアン検索モデル単独では，文書のランキングを行うことができない．よって本手法では，ブーリアン検索式の考え方と確率型検索モデルと組み合わせることにより，文書のランキングを行う．2.2 節で述べたブーリアン検索式の修正手法により，ユーザが与えた初期ブーリアン検索式を修正することが可能になったが，すべての適合文書集合と比較を行っているわけではないので，作成したブーリアン検索式が，完全にすべての適合文書を見つけることができるブーリアン検索式であるという保証がない．

よって，本手法では，ブーリアン検索式を満たさない適合文書がある可能性を考慮したうえで，次の 2 つの確率型検索モデルとの組合せ手法を提案する．

- ブーリアン検索式を適合文書判定のためのフィルタとして利用
  確率型検索モデルによる検索を実行し，スコアによるランク付きの検索結果を得る．この結果に対し，ブーリアン検索式を満たすかどうかをチェックし，ブーリアン検索式を満たさない文書については，そのスコアにかかわらず，検索結果から除外する．
- ブーリアン検索式を満たさない文書にペナルティを付与
  確率型検索モデルによる検索を実行し，スコアによるランク付きの検索結果を得る．この結果に対し，ブーリアン検索式を満たすかどうかをチェックし，ブーリアン検索式を満たさない文書については，ブーリアン検索式との違いに応じたペナルティを与え，スコアを下げ，ランキングをやり直す．このペナルティについては，各検索式における検索語の重要度に応じた値を設定する．

本手法で作成したブーリアン検索式が，適合文書のすべてを網羅できる形で作成できる場合には，ブーリアン検索式を満たさない文書を除外する前者の方法が有効であると考えられる．一方，作成したブーリアン検索式が不十分である場合には，前者の方法では，正解文書の一部を完全に取り除くことになり，後者の方法が（特に，再現率を重視する場合に）有効になると考えられる．

2.3　ABRIR

ABRIR では，Web のテキストといった大規模なテキストデータベースを利用するための検索エンジンとして，汎用連想計算エンジン GETA [⋆1]を利用した．また，本システムでは，形態素解析器として茶筅[7] を利用し，主に名詞を抽出し，インデックス語とした．さらに，名詞が連接している場合には，連接した 2 つの語を組み合わせたフレーズ語[8] を作成し，インデックス語とした．情報検索の確率型検索モデルとしては，Okapi BM25 [9] を利用した．Okapi BM25 においては，文書 d のスコア $S(d)$ を式 (1) のように定義する．

$$S(d) = \sum_{T \in Q} w^{(1)} \frac{(k_1+1)tf}{K+tf} \frac{(k_3+1)qtf}{k_3+qtf} \tag{1}$$

ただし，$w^{(1)}$ は検索語 T の重みで，以下の Roberson/Sparck Jones の式により求められる．

$$w^{(1)} = \log \frac{(r+0.5)/(R-r+0.5)}{(n-r+0.5)/(N-n-R+r+0.5)} \tag{2}$$

Q は検索式に含まれるすべての検索語の集合，N はデータベース中のすべての文書数，n は検索語 T を含むすべての文書数，R は与えた適合文書数，r は適合文書中で T を含む文書数，tf と qtf は T の文書中と検索式中での頻度，k_1，k_3，K は制御用のパラメータである．本手法では，$k_1 = 1$，$K = \frac{dl}{avdl}$，$k_3 = 7$ を利用した（dl は検索文書中に含まれる検索語の延べ総数，$avdl$ はすべての文書における dl の平均を示す）．

この式において，$w^{(1)} \frac{(k_3+1)qtf}{k_3+qtf}$ の部分が，検索式における検索語の重要度の値を示す部分であると考え，パラメータ β を利用し，ブーリアン検索式を満たさない場合のペナルティとして，以下の式を利用した．

$$Penalty(T) = \beta * w^{(1)} \frac{(k_3+1)qtf}{k_3+qtf} \tag{3}$$

また，or 式を満たさない場合には，or 式に含まれる語のうち，最も重要だと考えられる語の値を用いてペナルティとした．

n 個の or 式を and でつないだ形 $(T_{11}$ or $T_{12} \cdots$ or $T_{1o_1})$ and $(T_{21}$ or $T_{22} \cdots$ or $T_{2o_2})$ and $\cdots (T_{n1}$ or $T_{n2} \cdots$ or $T_{no_n})$ で表されるとする．ただし，各 or 要素で，Penalty は，降順に並んでいることとする（$Penalty(T_{i1}) \geq Peanlty(T_{12}) \geq \cdots Penalty(T_{io_i})$）．このとき，検索質問 Q に対するスコアの式 $S'(d)$ は，ブーリアン式の各々の or 式のうち，満たさ

⋆1 http://geta.ex.nii.ac.jp/

© 2011 Information Processing Society of Japan

ないものの集合を NS として，次の式で計算される．

$$S'(d) = \sum_{T \in Q} w^{(1)} \frac{(k_1+1)tf}{K+tf} \frac{(k_3+1)qtf}{k_3+qtf} - \sum_{i \in NS} Penalty(T_{i1}) \quad (4)$$

このペナルティの計算を，先の図 1 の例に基づき，具体的に説明する．まず，最初に，検索式に含まれる語（“A,” “C,” “D”）のペナルティを式 (3) を用いて計算する．ここで，$Penalty(C) \geq Penalty(D)$ だと仮定する．このとき，T_{11} が A に，T_{21} が C に，T_{22} が D に，対応する．このとき，“A,” “C,” “D” のいずれも含まない文書は，$Penalty(A)+Penalty(C)$ を確率型検索モデルによるスコアから減算し，“D” を含むが “A,” “C” を含まない文書は，$Penalty(A)$ だけ減算を行う．

また，検索語拡張の際には，疑似適合文書との相互情報量の高い単語から 300 語を選択し，検索語に追加し，最終の検索式を作成する．

3. ABRIR の質問応答への応用

3.1 Web 文書検索と質問応答のための情報検索の違い

前節で紹介した ABRIR は，主に，Web 文書検索のためにシステムが構築されていたため，そのままの設定では，質問応答のための情報検索システムには不適当であると考えられる．

まず，最初に，質問応答の質問と Web 検索の質問の違いについて述べる．質問応答では，「2002 年に合衆国がある国に侵攻したのは何月何日でしたか？」における「何月何日でしたか？」のように，解答のタイプを指定するのに役立つが，検索対象となる文章には含まれないと考えられる単語を含んだ形で質問が作成される．そのため，質問応答のための情報検索システムでは，その構文的な関係を使わない場合には，これらの単語は悪影響を与えるだけと考えられる．よって，これらの質問特有の表現を処理する必要がある．

また，Web 検索と質問応答のための情報検索では，文書を適合と判断する範囲が異なる場合が存在する．たとえば，NTCIR-4 の Web 検索タスクのテストコレクションのサーベイ検索[10] における検索実験では，ユーザが指定したブーリアン検索式を ABRIR を使って修正した場合においても，ブーリアン検索式を満たさない適合文書が数多く存在することが示されている[5]．これは，たとえば，作家の情報を検索する場合に，その作家が書いた本に関する情報などが含まれる場合にも，有用な情報と考える場合があるからである．

これに対し，特にある条件に合致する人名，地名，組織名などを問う factoid 型[11] の質問に対する質問応答のための情報検索では，その答えが記述されていることが必須であり，質問中に含まれている固有名詞や質問に関連する動詞の同義語などをすべて含んでいることが，より重要となると考えられる．そのため，適切なブーリアン式を作成することは，検索性能の向上につながると考えた．Web 検索においても，このように，少数のキーワードと関係性の深い関連文書を探し出す Target 型の検索[10] が存在し，そのような場合には，本研究と同様の手法が有用な場合が想定される．しかし，このような Web 検索に対しては，検索キーワードのタイプをアンカテキストを用いて分類し利用する方法[12] など，Web 特有の情報を利用する手法などがを適用する方が提案されているため，本手法は，主に，質問応答に用いる方法として提案する．

ただし，質問応答のための情報検索においても，ある事象が起きた理由を問うような複雑な質問[11] については，必ずしも，この方法が適切であるとは限らない．この質問タイプに応じて適切な設定を考えるという点については，4.3 節で議論を行う．

以下では，上記の情報検索の違いをふまえたうえで，主に factoid 型の質問を対象とした質問応答のための情報検索システムの構築方法の方針を述べる．

(1) 動詞を検索インデックスとする必要性
「A さんが亡くなったのはいつですか？」といった質問を扱うためには，動詞の情報をインデックス語に追加することが不可欠である．また，動詞は多くの同義語を持ち，これをうまく扱わないと，質問に対応する文書を見つけられないという問題が起こりやすくなると考えられる．

(2) 固有名詞（名付き実体：Named Entity）の取扱い
固有名詞に関する質問に対する文書には，固有名詞に関する情報が含まれていることが不可欠であるという考えを適用するためには，固有名詞を認識する技術が必要である．また，特に，外来語から派生したカタカナで記述される固有名詞については，表記のバリエーションについて，その対応方法を検討する必要がある．

(3) 適合文書の数
本課題における適合文書は，所定の事実に関しての記述を完全に含んでいる必要があるため，一般の Web 検索課題に比べて，1 つの事柄についての適合文書が少ない可能性が考えられる．このような課題に対して多くの疑似適合文書を用いることは，元のトピックからずれた検索要求を作る危険性が存在する．よって，疑似適合文書フィードバックなどを行う際の文書数については，Web 検索に比較して，少ない文書を使った方が，元の検索意図から大きくずれることがないと考えた．

© 2011 Information Processing Society of Japan

(4)　検索語拡張の語数

質問応答のための情報検索システムでは，再現率よりも精度を優先したい場合が多い．そのため，検索語拡張の語数は，少なくした方が良いと考えた．

これらの仮定について，その妥当性を実験により検証する．

3.2　質問応答用 ABRIR

検索用インデックスの作成は，Web 文書用のものとほぼ同じものを利用するが，動詞を扱うために，動詞をインデックスとして追加する．このインデックスに対して，以下の手順で検索式を作成する（**図 2**）．

(1)　質問文に特有の表現を削除

“のはいつですか？” といった質問内容に関係のない質問文に特有の表現を削除する．

(2)　形態素解析と固有名詞のタグ付け

Web 文書用とほぼ同じものを利用するが，インデキシング時と同様に，動詞も抽出する．また，固有名詞の抽出のために，Cabocha [13] を利用し，人名・組織名・地名などを固有名詞の候補として抽出する．

(3)　同義語リストの作成

抽出した動詞に対して EDR 電子化辞書[14] を利用して同義語リストを作成する．具体的には，1 つの動詞に対して，対応する 1 つまたは複数の意味 ID を調べ，その意味 ID に対応する語を同義語の候補とした．

(4)　カタカナの表記ぶれリストの作成

以下に示す書き換え規則を利用して同義語リストを作成する．

- オリジナルの表記から「ー」を削除する．
- オリジナルの表記から小さいカタカナ（例：“ｱｨｳｴｵﾔﾕﾖﾜｶｹﾂ”）を削除する．
- 小さいカタカナ（e.g., “ｱｨｳｴｵﾔﾕﾖﾜｶｹﾂ”）を大きいカタカナ（e.g., “アイウエオヤユヨワカケツ”）に置き換える[⋆1]．

この結果，「ヘップバーン」からは，「ヘップバン」，「ヘプバーン」「ヘツプバーン」の 3 種類の候補となる．

(5)　初期検索

確率型検索モデルを利用して，疑似適合文書を検索する．3.1 節に述べたように，適合文書の数が少ない課題が存在する可能性を考慮し，今回は，上位 3 件の文書を疑似

⋆1 大きいカタカナを小さいカタカナに置き換えるのは，無駄な表記を作成する可能性が高いので利用しなかった．

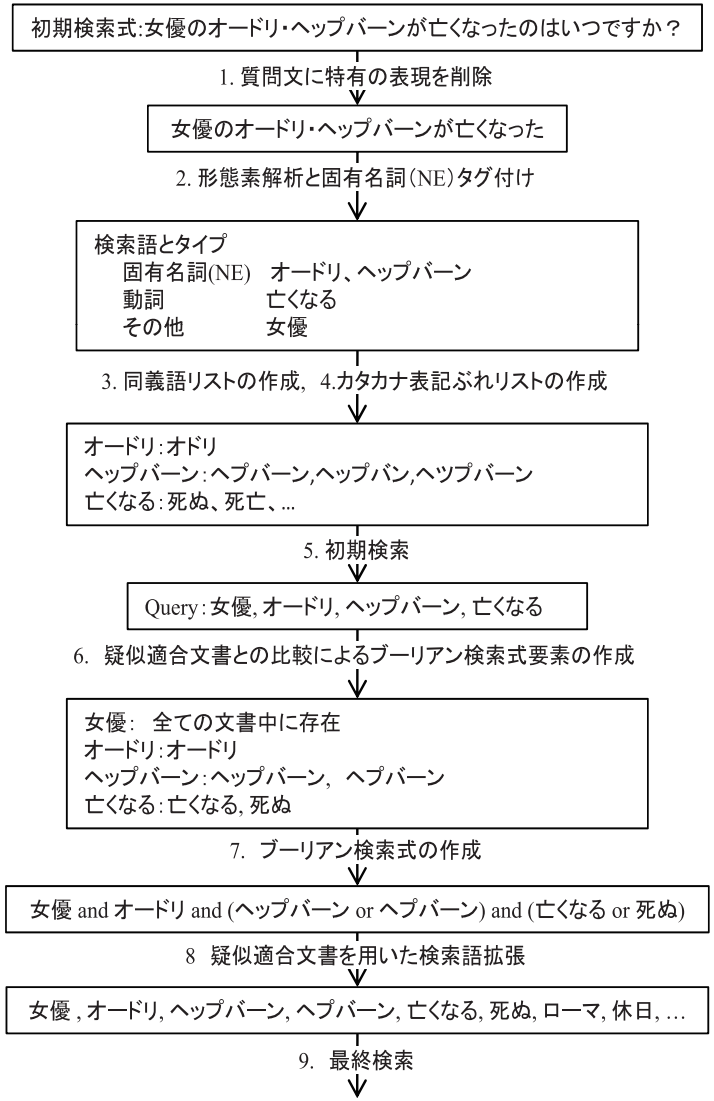

図 2　ABRIR での検索式の生成過程
Fig. 2　Procedures for query construction and retrieval in ABRIR.

© 2011 Information Processing Society of Japan

適合文書とする．

(6) 疑似適合文書との比較によるブーリアン検索式要素の作成
検索式中のキーワードは，固有名詞，動詞，その他に分類される．各々について，次のような基準でブーリアン検索式を作成する．

- 固有名詞
  システムが生成した固有名詞のバリエーションリストの多くは，実際の文書に現れないデータであり，無意味であるだけでなく，場合によっては，別の意味を持つ単語となってしまう場合が考えられる．この問題を避けるために，システムは，実際に，疑似適合文書に現れるものから，候補を選択し，異表記候補とする．たとえば，先ほどの「ヘップバーン」の例で，2 つのドキュメントには，「ヘップバーン」があり，もう 1 つのドキュメントには，「ヘプバーン」がある場合には，「ヘップバーン」or「ヘプバーン」というブーリアン検索式を作成する．
- 動詞
  動詞についても同様に，実際の疑似適合文書と比較して，同義語の候補を生成する．たとえば，「亡くなる」という検索語に対して生成したリストのうち，「死ぬ」という単語が適合文書中に存在した場合には，「亡くなる」or「死ぬ」というブーリアン検索式を生成する．
  動詞は，固有名詞に比べ，同義語が使われる可能性が高いため，質問対象の固有名詞に関する文書が多い場合に，注目している動詞に関係する同義語が疑似適合文書内に見つからない可能性が高い．そこで，すべての疑似適合文書中に，同義語が見つけられない場合には，すべての同義語を検索式に追加した追加検索を行い，新たな疑似適合文書を得る．この疑似適合文書について，同様の処理を行うことにより，新たな動詞に関する検索式を生成する．
  ただし，動詞に関する検索拡張を行った結果得られた疑似適合文書では，動詞に関する検索語の重みが大きすぎて，以降の処理を行うのに不適切であると考えられる．よって，実際の検索式に追加した動詞のみを追加した検索式を作成して，再検索を行うことによって，以降で用いる疑似適合文書とする．
- それ以外の単語
  固有名詞の判定間違いを行う可能性があるので，カタカナがある場合には，固有名詞と同様の処理を行う．また，疑似適合文書にすべて含まれる単語は，ABRIR と同様に，ブーリアン検索式に and として組み入れる．
  一方，今回の質問応答のための情報検索の課題のように精度重視という立場に立つと，疑似適合文書のすべてに存在しなくても，疑似適合文書中に含まれるのであれば，ブーリアン検索式の and として組み入れるという考え方もある．

(7) ブーリアン検索式の作成
上記の 3 種類の単語各々について作成したブーリアン検索式を and でつなぐことによってブーリアン検索式を作成する．

(8) 疑似適合文書を用いた検索語拡張
ABRIR と同様に，疑似適合文書中で高い相互情報量を持つ単語を検索語拡張の候補として追加する．ただし，3.1 節の議論をふまえ，今回は 5 単語のみを追加することとした．また，これに加え，ブーリアン検索式に追加された新たな単語はすべて，検索語に追加した．

(9) 最終検索
検索語拡張した検索式に基づき，確率型検索モデルを用いて情報検索を行う．この際に，以降のペナルティ計算で順位が逆転する可能性があることを考慮して，獲得したい検索結果の 1.5 倍（1,000 件の問合せの場合には，1,500 件）の検索結果を返す．この結果に対し，ブーリアン検索式と照らしあわせて，ブーリアン検索式とマッチしない部分についてペナルティを与える．ただし，固有名詞が重要であると考え，固有名詞に対するペナルティは，式 (3) を用いるのではなく，$Penalty(T) = 1{,}000{,}000$ とした．

4. 評価実験

4.1 実験データ

本システムの有効性を検証するために，NTCIR-8 GeoTime タスク[15] の日本語単言語検索に参加し，その性能を評価した．本タスクは，毎日新聞の 2002〜2005 年の記事データ（377,941 件）に対し，24 件の地理情報や時間情報に関する主に factoid 型の質問を与え，質問の答えを含む記事を見つけることが求められている．**図 3** に示す質問例にあるように，特定の人物に関する事象の起きた時間を問う質問や，位置情報を考慮した質問などが含まれる．すべての参加者は，各々の質問に対する検索結果としては，最大 1,000 件の結果を返すことができる．

本実験で用いたパラメータは，Web 文書検索[5] で用いたパラメータと同じものを利用した．具体的には，式 (1)，(2) におけるパラメータとして，$k_1 = 1$，$k_3 = 7$，$K = \frac{dl}{avdl}$，

© 2011 Information Processing Society of Japan

- いつ，どこでアストリッド・リンドグレーンは亡くなりましたか？
- いつ ITER の設置とその建設予定地が決定しましたか？
- 2002 年から 2005 年の間で，いつ，どこで，死傷者がでた列車事故が起こったかを述べよ．

図 3　NTCIR-8 GeoTime タスクにおける質問例
Fig. 3　Query example of NTCIR-8 GeoTime task.

$c = 0.3$，$\alpha = 0.7$ を利用した．ここで，dl は各文書の長さ（インデックス語の延べ語数）で，$avdl$ は，すべての文書における文書の長さの平均である．

ペナルティの式（式 (3)）の計算においては，$\beta = 3$ とした．

これらの共通設定のうえに，次のような異なる設定を行った実験を行うことにより，本システムの特徴を分析した．

NE-penalty-Other-all-penalty（NpOp）　基本設定：すべての単語について，ブーリアン検索式を満たさないものは，ペナルティ計算を行う．また，その他の単語については，疑似適合文書中のすべての文書に含まれる場合にブーリアン検索式に追加した．

NE-penalty-Other-all-penalty-NoV（NpOpNoV）　NpOp の設定を動詞をインデックス語として追加せずに作成した検索インデックスに適用．

NE-penalty-Other-all-penalty-Expansion（NpOpE30，NoOpE300）　NpOp の設定で，検索語拡張の語数を 30（NpOpE30），300（NpOpE300）に設定したシステム

NE-penalty-Other-all-penalty-Relevant-Expansion（NpOpR5，NpOpR5E300）　NpOp の設定で，疑似適合文書の数を 5 件（検索拡張語数は 5）にしたものと，疑似適合文書の数を 5 件にして，検索語拡張の語数を 300 に設定したシステム

NE-filter-Other-all-penalty（NfOp）　質問の対象となっている固有名詞に関係しない文章をフィルタリングするために，基本設定から，ブーリアン検索式の固有名詞相当部分（図 2 の場合は，オードリ and（ヘップバーン or ヘプバーン））を満たさないものは検索結果から除外した．

NE-filter-Other-all-penalty-Nov（NfOpNoV）　上記の設定を動詞をインデックス語として追加せずに作成した検索インデックスに適用．

ABRIR-Web　従来の ABRIR に相当するシステム．今回，提案したカタカナのバリエーションや動詞の同義語を用いたブーリアン式の修正は行わない．また，固有名詞を特別扱いしないで，すべて式 (3) を用いてブーリアン検索式のペナルティを計算した．さらに，以前のシステムと同様に，疑似適合文書数 5，検索語拡張は 300 語に設定した．

All-penalty-NoKatakanaExpansion（ApNoVENoK）　上記と同じ設定で，疑似適合文書数 3，検索語拡張は 5 語に設定した．

All-penalty-NoVerbExpansion（ApNoVE）　上記の設定に加え，カタカナのバリエーションによるブーリアン式の修正は行う．

All-penalty（Ap）　すべて式 (3) を用いてブーリアン検索式のペナルティを計算した．

Baseline-Okapi（BO）　提案システムにおいて，ペナルティの計算を行わない手法．提案システムの手順と同様に，質問文に特有の表現を削除した検索質問を利用した Okapi 単独での検索に相当．疑似適合文書を 3 件とし，検索語拡張では，他の手法が，固有名詞，動詞に関する検索語の追加を行っていることを考慮して，検索語拡張の語数は 10 とした．

4.2　実験結果に関する考察

表 1 に description（図 3 に示した，質問文による検索）のみを用いて検索した場合の，すべてのシステムの評価結果を示す．評価指標は，文献 15) と同様に，精度と再現率のバランスを考慮した平均精度（AP），平均精度に多値適合判定データ（適合・部分適合などを用いて関連文書の関連度を示す）を考慮した指標である Q-measure（Q），重要な正解をより上位で見つけた場合に値が高くなる正規化減損累積利得（nDCG：normalized Discounted Cumulative Gain）を利用した．この結果，基本設定 NpOp は，NTCIR-8 の GeoTime タスクで最も良い評価値を実現した．

まず最初に，NpOp と NpOpNov，ならびに，NfOp と NfOpNov の比較から，動詞をインデックス語に追加する効果を検証することができる．NpOp と NpOpNov の比較からは，11 件の質問に対し，評価値が改善され，8 件の質問に対して，評価値が低下した．平均的には，動詞をインデックスに追加する方が良い評価値になっているが，統計的には有意ではない．評価値が向上した質問には，「（人が）亡くなる」「（流行が）始まる」といった適合文書に多く含まれる動詞を含んだ質問であり，評価値が低下した課題には，「（事故が）起きる」といった比較的適合文書に，対応する動詞が存在しない場合がある質問であった．「（人が）亡くなる」「（流行が）始まる」といった動詞については，その動詞の情報を利用できないことは，大きな問題であり，特定のタイプの質問に対しては，動詞をインデックス語に使う効果が確認された．

次に，これまでの ABRIR-Web を用いた場合の検索性能ならびに，いくつかの改善案を

© 2011 Information Processing Society of Japan

表 1　異なる設定のシステムの評価結果（GeoTime）：太字が最大値
Table 1　Evaluation measures for each version of the system (GeoTime): The bold face number indicates the maximum value.

	AP	Q	nDCG
NpOp	0.379	0.424	**0.593**
NpOpNov	0.353	0.393	0.558
NpOpE30	0.381	**0.427**	0.591
NpOpE300	0.384	0.419	0.584
NpOpR5	0.373	0.421	0.584
NpOpR5E300	**0.388**	0.421	0.590
NfOp	0.372	0.411	0.563
NfOpNov	0.348	0.384	0.539
ABRIR-Web	0.334	0.379	0.543
ApNoVENoK	0.349	0.383	0.553
ApNoVE	0.372	0.400	0.570
Ap	0.366	0.409	0.577
BO	0.302	0.343	0.508

追加した場合の検索性能について議論を行う．

動詞をインデックス語に使う効果については，NpOp と NpOpNov の差などより，明らかであるため，その比較は割愛した．まず，ABRIR-Web と ApNoVENoK の差は，検索語拡張の語数と疑似適合文書の数である．この比較では，ApNoVENoK の方が良い評価値になっているが，検索語拡張の語数と疑似適合文書については，その他の機能との相互関係もあるので，後ほど，議論をすることとする．

次に，ApNoVENoK と ApNoVE を比較によるカタカナ語のバリエーションを用いてブーリアン検索式を作成する効果であるが，ブーリアン式の修正を行った 3 つの課題ですべて評価値が向上しており，本手法の有効性が確認された．

さらに，ApNoVE と Ap の比較から，動詞の同義語を利用したブーリアン式の作成の妥当性について考える．今回の実験では，5 件の検索課題について，ブーリアン式の修正が行われ，3 件では，質問の中心となる概念に関する動詞の同義語（「亡くなる」に対して「死去」）を見つけた場合には，評価値の改善が見られたが，「起きる」「生じる」といった必ずしも，中心的な概念でない動詞に対して，ブーリアン式を作った場合にやや評価値を落とすこととなった．結果として，上位でのより適切な適合文書に注目する指標である nDCG や Q は向上したものの，AP が低下するという結果となったと考えられる．

NpOp と NfOp の比較から，固有名詞でのみ構成されたブーリアン検索式の有効性が検証できると考える．今回の質問のうち，9 つの質問については（質問番号：4，10，11，16，18，21，22，23，24），絞り込みに役立つブーリアン検索式は作成されずに，1,000 件の結果を得ることとなった．しかし，11 件については，1,000 件より少ない検索結果で，すべての適合文書を見つけ出すことができた（質問番号：検索結果数 1：8，3：1，5：16，6：277，7：1，8：432，9：6，13：105，15：109，19：329，20：947）．ただし，4 件の検索課題では，不適切な絞り込みをしてしまったために，これらの文書を検索結果に含めることができなかった（質問番号：除外した適合文書数 2：27/48，12：15/19，14：2/2，25：1/3）．これに対し，ペナルティの計算を行う NpOp では，1,000 件までの結果を返すことで，これらの文書を見つけることが可能であった．結果として，この 4 件での評価値の差が全体の評価値の差となっている．

さらに，固有名詞を抽出することの有効性を検証するために，NpOp と Ap の比較を行った．この結果，9 件（1，5，6，8，12，13，15，23，25）の質問で NpOp の方が評価値が向上した．一方，2 件（2，14）の質問で，Ap の方が評価値が良かったため，その原因を検討する．

質問 2 は，「いつ，どこで，ハリケーン・カトリーナは，米国に上陸しましたか？」という質問であったが，固有名詞として，「ハリケーン・カトリーナ」を抽出した．これに対応するインデックスが「ハリケーン」＋「カトリーナ」であったため，「ハリケーン」と「カトリーナ」を両方含む文書を上位にランキングすることには成功したが，このハリケーンを「カトリーナ」という名前だけで参照し，「ハリケーン」という単語を含まない記事と，別の「ハリケーン」で「カトリーナ」を含まない記事が同じようなスコアとなり結果として，全体の評価値の低下をまねいたと考えられる．固有名詞抽出システムの性能の向上や，「カトリーナ」は「ハリケーン・カトリーナ」の略称であると行った対応関係を扱うことで，改善が期待できる．また，質問 14 は，「2002 年のアフリカで，いつ，どこで火山の噴火がありましたか？」という質問であるが，正解の文書には，「アフリカ」という単語が存在せず，「コンゴ民主共和国」から，「アフリカ」を対応づける必要がある課題であった．このような問題は，本研究で仮定した検索質問の想定外の問題であった．今後は，このような全体部分関係のような関係を扱う必要性があると考えている．

次に，NpOp，NpOpE30，NpOpE300，NpOpR5，NpOpR5E300 を比較することにより，検索語拡張の語数と利用する疑似適合文書の数について議論する．本システムでは，ABRIR-Web の場合と異なり，検索語拡張の語数の追加しても，評価値の悪化が見られなかった．これは，重要な固有名詞については，それを含まないことのペナルティが存在する

© 2011 Information Processing Society of Japan

ために，多数の検索語拡張を行った場合でも，中心的なトピックから，大きく外れることがなかったためではないかと考えている．一方，これらの実験結果の課題ごとの比較を行ったところ，約半数で評価値が向上し，約半数で評価値が低下するといった状況であり，質問のタイプなどを考慮して，より詳細な検討を進める必要があると考えている．

最後に，今回のシステムの中で最も評価値の高かった NpOp とベースラインシステムを比較する．この結果，20 件の質問に対しては，評価値が向上し，4 件の質問に対しては，評価指標の一部で評価値の低下があった（2：nDCG，Q，5：nDCG，11：AP，nDCG，Q，21：AP）．2 については，先ほど述べた固有名詞の抽出の失敗の影響が大きいと考えられる．また，5 については，「キャサリン・ヘップバーンはいつ亡くなりましたか」という課題に対して，「キャサリン・ヘップバーンが亡くなった」ことは書いてあるが，時期は書いていないという記事の順位が悪影響を与えただけであり，その影響もほとんどない．

11 は，「2002 年から 2005 年の間で，いつ，どこで，死傷者がでた列車事故が起こったかを述べよ.」という質問であるが，正解文書 253 件に対して，ブーリアン検索式の大部分を満たすような文書が少なく，28 件目からスコアが負の値になってしまった．このようなサーベイ型の質問に対し，厳しいブーリアン検索式を作った場合には，全体のパフォーマンスの低下をまねくので，質問タイプの推定などと組み合わせた分析などを行うことが望まれる．21 については，AP をのぞいては，提案システムの方が評価値が良く，大きな問題ではないと考えている．

また，提案システム（NpOp）とベースライン（BO）の各評価値に対し，統計的検定を行った．まず，AP，Q，nDCG の 3 つの指標について危険率を 5%とした正規性の検定を行ったところ，AP についてのみ正規性の仮説が棄却された．これをふまえ，Q と nDCG については，危険率を 5%とした t 検定を行い，AP については，同じく危険率を 5%とした Wilcoxon の符合付き順位和検定を行った．この結果，各々の指標の差が統計的に有意であると判断された．

4.3　他のテストコレクションによる検証

さらに，本システムの有効性を検証するために，より課題数の多い NTCIR-8 の質問応答タスクである ACLIA の質問応答のための情報検索を評価するテストコレクションを利用し評価を行った．本テストコレクションは，GeoTime と同様に毎日新聞の 2002〜2005 年の記事データ（377,941 件）に対する 94 件の質問から構成されている．

ACLIA には 94 問の質問中に PERSON，ORGANIZATION，LOCATION，DATE に関する factoid 型の質問が約 2 割，DEFINITION，BIOGRAPHY，RELATIONSHIP，EVENT，WHY に関する複雑な質問が約 8 割存在する．しかし，質問ごとにどの質問タイプに相当するかという記述は存在しないため，質問の内容から，factoid 型の質問 18 問と複雑な質問 76 問に分類した．

すべての課題についてのシステムの評価値と Factoid 型質問に対する評価値を**表 2**，**表 3** に示す．

この結果，GeoTime の場合と異なり，固有名詞を特別扱いしない Ap の評価値が NpOp

表 2　異なる設定のシステムの評価結果（ACLIA）：太字が最大値
Table 2　Evaluation measures for each version of the system (ACLIA): The bold face number indicates the maximum value.

	AP	Q	nDCG
NpOp	0.447	0.472	0.686
NpOpNov	0.459	0.485	0.703
NpOpE30	0.456	0.481	0.690
NpOpE300	0.471	0.497	0.700
NpOpR5	0.444	0.470	0.686
NpOpR5E300	0.473	0.499	0.704
NfOp	0.400	0.417	0.609
NfOpNov	0.414	0.430	0.623
ABRIR-Web	**0.483**	**0.509**	**0.709**
ApNoVENoK	0.454	0.481	0.690
ApNoVE	0.454	0.481	0.690
Ap	0.452	0.478	0.688
BO	0.428	0.452	0.663

表 3　異なる設定のシステムの評価結果（ACLIA：Factoid 型質問）：太字が最大値
Table 3　Evaluation measures for each version of the system (ACLIA): The bold face number indicates the maximum value.

	AP	Q	nDCG
NpOp	0.411	0.465	0.686
NpOpNov	0.403	0.442	0.678
NpOpE300	0.424	**0.486**	**0.701**
NpOpR5E300	0.414	0.475	0.692
ABRIR-Web	0.419	0.485	0.692
ApNoVENoK	**0.425**	0.480	0.691
ApNoVE	**0.425**	0.480	0.691
Ap	0.424	0.479	0.691
BO	0.388	0.454	0.665

© 2011 Information Processing Society of Japan

の評価値を上回った．また，多くの検索語拡張を行う ABRIR-Web の評価値が最も良い結果となった．また，factoid 型の質問のみを対象とした評価値に比べて，すべての質問を使った評価値の差が大きいことが判明した．このことから，複雑な質問が多い場合には，検索語数を多くすることが適切な場合があることが示唆される．

また，動詞をインデックス語に追加する効果についても，factoid 型とその他の質問では性質が異なった．これは，factoid 型では，質問の中心となる動詞の存在が重要な場合が多いのに対し，複雑な質問の例である RELATION では，「劉翔とジョンソンの関係は？」のように，動詞を必要としない質問が多く存在するのが 1 つの原因と考えられる．また，質問のバリエーションが増えたために，質問特有のタイプを表す表現が十分に削除できず，不要な動詞を追加してしまったケースも存在した．

ACLIA のカタカナ語のバリエーションに対しては，ApNoVENoK と ApNoVE の差がないように，有効なバリエーションが生成されず，検索結果が変わらない結果となった．GeoTime では，最も有効に作用したこのようなバリエーションに対応する課題が存在しなかったことが，本提案手法の評価値が ABRIR-Web について低くなった大きな要因であると考えている．また，動詞のバリエーションについては，GeoTime のときとほぼ同じ傾向を示した．

最後に，固有名詞を特別扱いする効果について確認するために，固有名詞に注目することにより，評価値が向上した課題と低下した課題について確認をした．評価値が向上している課題には，「劉翔とジョンソンの関係は？」といった明らかに，2 人の名前を含んでいないと正解とならないような課題があった．これに対し，評価値を大きく下げた課題には，「2003 年に逮捕されたロシアの石油大手ユコス社の元社長はだれですか.」に対し，固有名詞として，「石油大手ユコス社」を抽出し，「ユコス社」単独がこれにマッチすることを判断できなかった場合や，「イギリス」，「ロシア」，「アメリカ」など，新聞では，「英」，「露」，「米」などと表記される場合がある国の名前などを含む課題が多く見られた．GeoTime の課題と同様に，固有名詞の抽出と同一性の判定について，より工夫が必要であることが，こちらの課題からも判明した．

また，提案システム（NpOp）とベースライン（BO）の各評価値に対し，同じく統計的検定を行った．今回は，AP，Q，nDCG のすべての指標について危険率を 5%とした正規性の検定を行ったところ，すべての指標で正規性の仮説が棄却された．よって，すべての指標について危険率を 5%とした Wilcoxon の符合付順位和検定を行った．この結果，各々の指標の差が統計的に有意であると判断された．

4.4　今後の課題

NTCIR-8 GeoTime タスクのデータをもとに行った比較から，本システムにおいて最も効果があった修正は，固有名詞のバリエーションに注目したうえで，そのバリエーションを考慮したブーリアン式の作成を行う点にあると考えられる．

一方，動詞をインデックス語に追加する効果，検索語拡張の語数，利用する疑似適合文書の数については，質問タイプの影響などが大きいことが示唆されているため，今後，質問タイプとの関係について考察を進めていく必要があると考えている．

一方，「ハリケーン・カトリーナ」と「カトリーナ」の関係や，「アフリカ」と「コンゴ共和国」のように，単純なバリエーションでは扱えない関係が存在する場合に，固有名詞に関するペナルティが悪影響を及ぼすことも確認された．

これらの問題に対しては，固有名詞のバリエーションを単純なカタカナだけではなく，「米国」→「アメリカ」のような組合せを利用したり，「アフリカ」→「コンゴ共和国」のような全体部分関係を扱うような枠組みを利用したりすることにより，安定的にシステムの性能の向上をはかることができると考えている．

それ以外に提案をした動詞のバリエーションの利用，検索語拡張の語数の変更，利用する疑似適合文書の文数については，検索課題ごとの影響が大きいこと，上位の精度を向上させることと，全体の再現率をあげることのバランスなどを考慮して決める必要があり，検索課題のタイプの推定と組み合わせて検討する必要があると考えている．

5. 関 連 研 究

質問応答のための情報検索の研究としては，質問タイプに注目して検索質問を修正する方法[16] や，質問タイプから想定される答えを含む可能性の高い文書に注目してリランキングする方法[17),18] や，本システムで用いているようなブーリアン検索式を利用する方法[4] など，様々な方法が提案されている．

Agichtein ら[16] は，質問の答えを含む文書には，質問タイプ（5W1H など）に応じて，答えの記述形式に特徴があることに注目した検索語拡張方法を提案している．具体的には，質問タイプごとに，質問と答えを含む文章のペアから，答えを記述する際によく使われるフレーズを抽出する．このフレーズを検索語拡張の語として追加することにより，Web に対する情報検索の性能が向上することを確認している．本研究で提案している検索語拡張とは補完的な役割があると考えられるが，同じ，「どこですか（Where）？」に対しても，動詞が違うと，フレーズの形が異なるといった問題もあり，簡単に組み合わせることは困難である．

© 2011 Information Processing Society of Japan

Mori[17]は，1回目の検索で得られた文書集合から得られた解候補の情報を検索語拡張に用いる方法を提案している．また，Kishida[18]は，地理時間情報に関する質問応答のための情報検索においては，質問タイプに応じて，時間や場所に関する情報のバリエーションが多く記入されているものが解答を含む文書である可能性が高いことに注目して，たとえば，地理情報に関する質問では，場所に関する用語を数多く含むもののランクをあげるランキング手法を提案している．これらの2つのシステムでは，GeoTimeのテストコレクションに対する実験結果が報告されているが，本システムの評価値を下回る結果となっている．

Saggionら[4]は，質問応答のための情報検索システムにブーリアン検索を利用する際に，ブーリアン検索式を修正する方法を提案している．この研究では，初期の検索式をブーリアンのand検索で行い，その結果が0件のときに，本手法と同様に，WordNetをシソーラスとして用いて検索語拡張を行ったブーリアン検索式を作成し，ブーリアン検索を行う．本手法では，ブーリアン検索式で1件でも関連文書が見つかった場合に検索語拡張を行わない点や，作成したブーリアン検索式を満たさない関連文書中を調べないという点で異なる．また，本システムで最も効果のあった固有名詞表現のバリエーションについては取り扱っていない．

Shimaら[19]は，本システムと同様に固有名詞などの重要性に注目し，それらの単語の検索式における重要度を高くすることにより，より適切な文書を発見する方法を提案している．結果として，単語が存在しない場合に，スコアが高くならないという意味で，類似したアプローチであるが，存在しないということを積極的に評価していない点が異なる．また，ACLIAのテストコレクションに対する実験結果が報告されているが，本システムの評価値を下回る結果となっている．

本研究で注目している，文書中の表記とユーザの検索要求で使われる単語の間の違いがある[20]ことに注目した研究としては，クエリ書き換えの研究がある[21),22)]．

クエリ書き換えの研究では，主に，ユーザが与えた単語に誤りがある場合に意味的に類似している単語に置き換えるという研究である．Jonesら[21]や萩原ら[22]は，検索ログにおける単語の共起関係などから，同じ意味を持つと考えられる異表記の情報を見つけ出す方法を提案している．実際に使われている検索語の組合せに注目して異表記の候補を生成するという意味では，本システムが初期検索結果の中から異表記の候補を探し出すという点で共通しており，異表記バリエーションの生成に役立つと考えられるが，今回のように，or式を含むようなブーリアン検索式を構築するという考え方とは異なる．

6. おわりに

本研究では，質問応答のための情報検索システムを作成するにあたり，ブーリアン検索の考え方を取り入れることが性能の向上につながると考え，確率型検索モデルとブーリアン検索モデルを組み合わせた情報検索システムABRIRに基づいた情報検索システムを提案した．

また，本システムを，NTCIR-8 GeoTimeタスクに適用したところ，確率型検索モデルに基づく情報検索システムを利用する場合に比べて，統計的に有意な性能向上を果たすことが確認された．

一方，NTCIR-8 ACLIAのテストコレクションに適用した結果から，固有名詞の同一性の判定に関して適切な知識が不足している場合に，固有名詞を特別扱いする手法の方が全体の評価値が悪くなるという結果が得られている．この問題を解決するためには，固有名詞の抽出システムの性能改善や，地名の全体部分関係の知識などを含む，地名の表記バリエーションについての知識を整備し，利用する必要があると考えている．

謝辞　本研究の一部は，科研費基盤研究（B）21300029により行われた．また，NTCIR-8 GeoTime，ACLIAのテストコレクションを作成していただいたオーガナイザの皆様に謝意を記す．

参考文献

1) Greenwood, M.A. (Ed.): 22nd International Conference on Computational Linguistics, *Proc. 2nd Workshop on Information Retrieval for Question Answering* (2008), available from ⟨http://www.aclweb.org/anthology/W/W08/W08-18.pdf⟩.
2) Sakai, T., Kando, N., Lin, C.-J., Mitamura, T., Shima, H., Ji, D., Chen, K.-H. and Nyberg, E.: Overview of NTCIR-7 ACLIA IR4QA Task, *Proc. 7th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Quesiton Answering, And Cross-Lingual Information Access*, pp.63–93 (2010).
3) Sakai, T., Shima, H., Kando, N., Song, R., Lin, C.-J., Mitamura, T., Sugimoto, M. and Lee, C.-W.: Overview of NTCIR-8 ACLIA IR4QA, *Proc. 8th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Quesiton Answering, And Cross-Lingual Information Access*, pp.63–93 (2010).
4) Saggion, H., Gaizauskas, R., Hepple, M., Roberts, I. and Greenwood, M.A.: Exploring the Performance of Boolean Retrieval Strategies For Open Domain Question Answering, *Proc. IR4QA Workshop at SIGIR* (2004).

© 2011 Information Processing Society of Japan

5) Yoshioka, M. and Haraguchi, M.: On a Combination of Probabilistic and Boolean IR Models for WWW Document Retrieval, *ACM Trans. Asian Language Information Processing* (*TALIP*), Vol.4, pp.340–356 (2005).
6) Baeza-Yates, R. and Ribeiro-Neto, B.: *Modern Information Retrieval*, Addison-Wesley (1999).
7) 松本裕治，高岡一馬，浅原正幸：日本語形態素解析システム『茶筌』version 2.4.3 使用説明書，奈良先端科学技術大学院大学情報科学研究科自然言語処理学講座 (2008).
8) Toyoda, M., Kitsuregawa, M., Mano, H., Itoh, H. and Ogawa, Y.: University of Tokyo/RICOH at NTCIR-3 Web Retrieval Task, *Proc. 3rd NTCIR Workshop on Research in Information Retrieval, Automatic Text Summarization and Question Answering*, pp.31–38 (2002).
9) Robertson, S.E. and Walker, S.: Okapi/Keenbow at TREC-8, *Proc. TREC-8*, pp.151–162 (2000).
10) Eguchi, K., Oyama, K., Aizawa, A. and Ishikawa, H.: Overview of the Informational Retrieval Task at NTCIR-4 WEB, *Proc. 4th NTCIR Workshop on Research in Information Access Technologies Information Retrieval, Question Answering and Summarization* (2004), available from ⟨http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings4/WEB/NTCIR4-OV-WEB-A-EguchiK.pdf⟩.
11) Mitamura, T., Shima, H., Sakai, T., Kando, N., Mori, T., Takeda, K., Lin, C.-Y., Song, R., Lin, C.-J. and Lee, C.-W.: Overview of NTCIR-8 ACLIA Tasks: Advanced Cross-Lingual Information Access, *Proc. 8th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Quesiton Answering, And Cross-Lingual Information Access*, pp.15–24 (2010).
12) 藤井　敦：アンカーテキストモデルと検索質問分類による Web 文書検索の高度化，情報処理学会論文誌，Vol.51, No.12, pp.2330–2342 (2010).
13) Kudo, T. and Matsumoto, Y.: Japanese Dependency Analysis using Cascaded Chunking, *CoNLL 2002: Proc. 6th Conference on Natural Language Learning 2002* (*COLING 2002 Post-Conference Workshops*), pp.63–69 (2002).
14) 日本電子化辞書研究所：EDR 電子化辞書（第 2 版）仕様説明書，TR2-006（改）(2001).
15) Gey, F., Larson, R., Kando, N., Machado-Fisher, J. and Sakai, T.: NTCIR-GeoTime Overview: Evaluating Geographic and Temporal Search, *Proc. 8th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Quesiton Answering, And Cross-Lingual Information Access*, pp.147–153 (2010).
16) Agichtein, E., Lawrence, S. and Gravano, L.: Learning search engine specific query transformations for question answering, *WWW '01: Proc. 10th International Conference on World Wide Web*, pp.169–178, ACM, New York, NY, USA (2001).
17) Mori, T.: A Method for GeoTime Information Retrieval based on Question Decomposition and Question Answering, *Proc. 8th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Quesiton Answering, And Cross-Lingual Information Access*, pp.167–172 (2010).
18) Kishida, K.: Vocabulary-based Re-ranking for Geographic and Temporal Searching at NTCIR GeoTime Task, *Proc. 8th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Quesiton Answering, And Cross-Lingual Information Access*, pp.181–184 (2010).
19) Shima, H. and Mitamura, T.: Bootstrap Pattern Learning for Open-Domain CLQA, *Proc. 8th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Quesiton Answering, And Cross-Lingual Information Access*, pp.37–42 (2010).
20) Custis, T. and Al-Kofahi, K.: A New Approach for Evaluating Query Expansion: Query-Document Term Mismatch, *SIGIR 2007: Proc. 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.575–582, ACM, New York, NY, USA (2007).
21) Jones, R., Rey, B., Madani, O. and Greiner, W.: Generating Query Substitutions, *WWW '06: Proc. 15th International Conference on World Wide Web*, pp.387–396 (2006).
22) 萩原正人，鈴木久美：意味的類似度を利用した日本語クエリ書き換え，言語処理学会第 15 回年次大会発表論文集，pp.522–525 (2009).

(平成 23 年 4 月 11 日受付)
(平成 23 年 9 月 12 日採録)

吉岡　真治（正会員）

昭和 44 年生．平成 8 年東京大学大学院工学系研究科精密機械工学専攻博士課程修了．同年より学術情報センター助手．平成 12 年より国立情報学研究所助手．平成 14 年より北海道大学助教授．現在，北海道大学大学院情報科学研究科准教授．情報検索への知識処理技術の応用，設計環境の知能化等の研究に従事．博士（工学）．ACM，人工知能学会，電子情報通信学会，言語処理学会各会員．

© 2011 Information Processing Society of Japan