地理的制約を加味した概要文生成を行う地理情報検索

安田　宜仁† 　戸田　浩之† 　松浦由美子† 　片岡　良治†

† 日本電信電話 (株), NTT サイバーソリューション研究所
E-mail: †{yasuda.n,toda.hiroyuki,matsuura.yumiko,kataoka.ryoji}@lab.ntt.co.jp

あらまし　地理情報検索とは, 内容を示すキーワードとユーザが関心のある地理的範囲という 2 つの制約を受け付けるような情報検索である. 地理情報検索は性質の異なる 2 種類の制約を受け付けるという特徴に加え, 地理的範囲は文字列によって表現できないため, ランキング方法だけでなく, 結果の提示においても新しい仕組みが必要となる. 我々が提案する地理情報検索システムは, 検索結果の概要文生成において, 性質の異なる 2 種類の制約を考慮し, 文字列のマッチングだけでなく, 地名間の連続的な距離と, 各地名の含意する広さを考慮するという特徴を持つ. 本稿では, 我々の提案する地理情報検索システムを紹介し, 提案する概要文生成の性能の評価結果を報告する.

キーワード

Norihito YASUDA†, Hiroyuki TODA†, Yumiko MATSUURA†, and Ryoji KATAOKA†

† NTT Cyber Solutions Laboratories, NTT Corporation
E-mail: †{yasuda.n,toda.hiroyuki,matsuura.yumiko,kataoka.ryoji}@lab.ntt.co.jp

Abstract　Geographic Information retrieval is a kind of information retrieval that accepts both keywords which reflect interested content and geographic constraints which reflect interested geographic area. Since geographic information retrieval system has special features such that it accepts two kinds of constraints as a query and geographic-part of the query is not represented by strings, it is required not only a new ranking method but also a new way of displaying search results. Our proposed geographic information retrieval system provides a novel summarization generation method that considers two kind of constraints; It considers continuous distances between geographic points and geographic extents implicated by each geographic names, besides usual string-matching based generation. In this paper, we introduce the proposed geographic information retrieval system and present the evaluation results of the proposed summarization generation method.

1. はじめに

情報検索技術, 特にインターネット上の膨大な文書からの検索技術の普及を背景に, 人々は日常生活のさまざまな判断の手助けに情報検索を用いるようになっている. また, 近年の携帯電話からのインターネット接続の普及に伴い, 特に外出時においては, その周辺ですぐに行う行動についての情報が必要とされることが多いと考えられ, 情報検索においてユーザの居場所に関する制約を適切に処理することは重要である.

地理情報検索とは, ユーザが意図する情報の内容とユーザが意図する地理的範囲という 2 つの制約によって情報を検索する情報検索である. 従来, 地理情報検索の主要な対象は, 「オーストラリアとカリフォルニアでの鮫の被害」といった比較的広い範囲を取り扱うものであった [1]. 一方, 我々は, 上述のようにたとえば外出先で携帯端末を用いて自身の居場所のごく周辺の情報を得るために用いることができる地理情報検索をめざしている. このため, ひとつの文書中に出現した複数の地名表現それぞれの含意する広さを考慮した検索手法を提案している [2]. 地名表現の含意する広さを考慮した重み付けを行うことにより, たとえば県名のように広い範囲を示す地名表現が, さまざまな地理的制約にマッチして, 結果として関連性の低い文書が検索結果が検索されてしまうという問題を防いでいる.

地理情報検索はキーワードによる一般的な情報検索とは異なる検索方式であり, ランキング手法だけでなく, システム全体の入力・出力を含めた新しいデザインが必要となる. 幸い, 近年の携帯端末は GPS を内蔵していたり, 基地局からの電波強度を用いて, 端末自身の場所を測位することが可能になっている. 本システムではこのような携帯端末自身が送信する位置情報と, ユーザが入力したクエリを入力と想定する.

地理情報検索の出力について, どのように結果を提示するのが適切であるかについては定まった考え方はない. 地理という特別な性質を活かし, 3 次元可視化 [3] や, 検索結果の文書を対

応する地図上の領域を Google Earth 上で提示すること [4] が提案されている.

一方, 一般的な情報検索においては, 各文書のタイトルと概要文を提示するという方法が広く用いられている. 地理情報検索も地理的制約の部分を除けば一般的な情報検索であるということ, 携帯端末等描画領域が小さい場合でも表示可能であることを踏まえて, 我々はタイトルと概要文による結果の提示方法を選択した.

しかし, 従来の概要文生成技術 [5], [6] はテキストによるクエリのみを想定しているため, 直接地理的制約を取り扱うことができない. そこで, 我々は地理情報検索におけるクエリの特性を踏まえた新しい概要文生成方法を提案する. 我々が考える地理情報検索におけるクエリの特性とは, まず, 内容を示すキーワードと, 地理制約という性質の異なる 2 種類の制約を受け付ける点にある. さらに, 地理的な距離や重なりは連続的なものなので, 地名同士がマッチするかどうかでは適切に地理情報を取り扱うことができない. また, 上述の通り, 各地名表現はそれぞれ含意する広さを持っている. このような特性を踏まえた新しい概要文生成を提案する.

本稿では, 我々の提案する地理情報検索システムを紹介し, 地理的制約を加味した概要文生成について述べる.

2. システム構成

本節では提案システムの構成について述べる.

システムは, 地理インデクス, 全文検索エンジン, 地名広さデータベース, ランキングモジュール, 概要文生成モジュールから構成される. システム構成図を図 1 に示す. 地理インデクスは, 文書中の地名表現の代表地点の座標を格納したものである. 地理インデクスの作成には, 平野らによる手法 [7] を利用し, 断片的な地名表現に対してもできる限り座標を付与している. 全文検索は地理インデックスを参照して, 検索対象文書を文書中の地名表現のうち少なくともひとつの座標が特定の範囲内にある文書に限定する機能を持つ. 文書のランキングは内容クエリを用いて, TFIDF に基づく一般的な方法によって行う [8]. 我々が提案する地理情報検索システムでは, ランキング, 概要文生成双方で, スコア付けにおいて地名の含意する広さを利用するため, 地名の広さを格納したデータベースを持つ. 地名広さデータベースは, 地名の先頭からの部分文字列 (接頭辞) に対して, その接頭辞を持つようなすべての地名の座標が構成する最小外接矩形の広さを対応付けたものである. ランキングモジュールは, 検索位置のごく周辺について述べた文書を優先的に上位のランクとするランキング方式である. 詳細は [2] を参照されたい. 概要文生成モジュールは, 従来から広く検索結果の概要文提示に用いられてきた query-biased 要約の手法に基づき, 地理的制約を加味をした概要文の生成を行う. 詳しくは 3. 節で述べる.

地理情報検索は一般に, 内容を示すキーワード (内容クエリ) とユーザが関心のある地理的範囲という 2 つの制約 (地理クエリ) を受け付ける. 本システムは地理クエリとして, 関心のある地理的範囲の中心点と, その中心点からの最大距離を与えることを想定している. このため, 本システムへのクエリは以下の 3

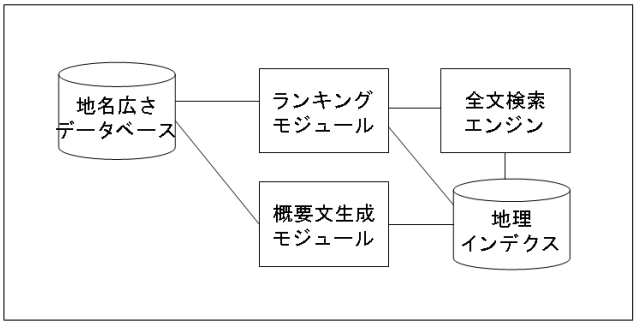


図 1　システム構成図

つの要素を含むことを想定している:
(1) キーワード (内容クエリ),
(2) 地理的範囲の中心点の座標, (3) 中心点からの最大距離.
このうち, 地理的範囲の中心点については典型的にはユーザの居場所を想定している. また, 最大距離については, 中心点からの直線距離を表すもので, ユーザの関心の範囲, あるいはユーザの移動手段などに応じて選択することを想定している.

クエリを受け取った後, システムは以下の手順で動作する.

1. まず, 内容クエリで示された検索語句をすべて含み, かつ, 地理クエリが示す範囲内に代表点を持つような地名表現を少なくともひとつ含む文書群を特定する.

2. 次に, 1. で特定した文書群を, 内容クエリを用いて TFIDF でのスコアリングを行う (内容スコア).

3. スコアリング結果の上位の文書 (実験では 2000 件) を選択する.

4. 3. の各文書について, 文書中の各地名表現と地理クエリの関係に基いたスコアを算出し, 文書の地理的スコアを算出する.

5. 地理的スコア, 内容スコアの積を文書スコアとして, ランキングを行う.

6. ランキング上位の文書 (ユーザ端末に出力する文書) について, 概要文生成モジュールを用いて概要文を生成し, 結果をユーザへ提示する.

3. 概要文生成

本節では概要文生成方法について述べる. 提案する概要文生成は, キーワードの密度に基くベースライン生成法に対して, 複数の制約を受け付けるための変更, および, 地理的制約を考慮するための変更を加えたものである.

3.1 基本生成法

ベースラインとなる概要文生成法として, 既存の概要文生成手法 [5], [6] を参考に, クエリ中の各キーワードを高密度に含むテキスト断片を結合したものを考える. これらの従来法は単語の頻度や位置といった表層的な処理によって利用可能な素性を用いて概要文を生成する方法である. 一方で, 深い言語処理を利用することで, より高精度な概要文を生成することをめざした手法がいくつか提案されている [9]〜[11]. しかし, 広く一般向けに公開するような検索システムで利用する場合, 処理コストは無視できない. このため, 我々は表層的な処理による概要文生成を基本手法とする.

この基本手法での概要文は以下に定めるテキスト断片 $F(l,x)$ のスコア $S(l,x)$ が高い断片によって構成する．

$$S(l,x)=\sum_{t\in T}\left(\alpha+(1-\alpha)\cdot(1-\frac{|mid(l,x)-kmid(t)|}{l/2})\right)(1)$$

ここで, T はクエリ中のキーワードの集合, x は文書中でのオフセット, l はテキスト断片の長さ, $mid(l,x)$ は $F(l,x)$ の中心位置, $kmid(t)$ は $F(l,x)$ 中に出現するキーワード t の中心の位置を示す．これらの変数の関係について図 2 に示す．$\alpha(0<\alpha<1)$ は, テキスト断片中でのキーワードの位置に基づくスコアの変動幅を調整するパラメータである．

基本手法では基本的にはこのスコアが高いテキスト断片によって構成するが, 何の制約もなくスコアだけを用いてテキスト断片を選択した場合, クエリ中のキーワードだけをテキスト断片として結合した概要文が選択されてしまう．しかし, そのような概要文がユーザにとって分かりやすいものではないことは自明であるのでテキスト断片の長さについて 2 つの対策を行う．ひとつはテキスト断片の最小の長さに制約を設けることである．これにより過度に短いテキスト断片で構成される概要文の生成を防ぐ．実験ではこの長さは 20 文字とした．もうひとつは, 多くのテキスト断片から構成される概要文の評価値を低減させる項を設けることである．これにより, より少ないテキスト断片で構成されるスニペットを優先する．

以上のテキスト断片のスコアと長さに関する制約を踏まえた基本手法による概要文の生成の手順は以下の通りである．

（1） あらかじめ決められた概要文の最大長 L および最小テキスト断片サイズ L_{min} を元に, 最大何個 (N_{max}) のテキスト断片からスニペットを生成するかを算出.

$$N_{max}=\lfloor\frac{L}{L_{min}}\rfloor$$

この値は 2 つのパラメータ $(L,\ L_{min})$ が決定していれば固定の値である．上述の通り, L_{min} は 20 とした.

（2） すべてのテキスト断片数 $N(N=1,..,N_{max})$ の場合について, スニペット候補を生成し, そのスコアを計算する．以下, それぞれのテキスト断片数ごとに行う処理の詳細を示す.

（a） テキスト断片のサイズ $L(N)$ を算出する.

$$L(N)=\lfloor L/N\rfloor$$

（b） すべてのテキスト断片 $F(L(N),x)$ について, (1) 式に基づきスコア $S(L(N),x)$ を算出する．ここで x は, 文書中でのテキスト断片の開始位置を示す.

（c） スコア $S(L(N),x)$ の高い順に N 個のテキスト断片を集める．ここで $X(N)$ はこのテキスト断片の集合を表す．この時, 既に集められたテキスト断片と位置的に重なりがあるテキスト断片は除外する.

（d） 上記で選択したテキスト断片の集合 $X(N)$ から, 一つのスニペットを生成し, 概要文のスコアを以下の式で算出する.

$$S(L(N))=\sum_{x\in X(N)}S(L(N),x)\cdot D^N$$

D は, 多くのテキスト断片から生成される概要文のスコアを低く見積もるための項である $(0<D<1)$.

（3） 前のステップで生成した概要文候補のうち最もスコアの高いスニペットを生成した条件 (N) を特定し, その条件で生成された概要文を出力する.

3.2 2 種類の制約を考慮するための生成法

地理情報検索では, 内容と地理という 2 種類の制約を受け付ける．このような性質の異なる 2 種類の制約に対応する概要文生成方法について述べる．このような 2 種類の制約を受け付ける場合, どちらか一方の制約のみを反映した概要文ではなく, 双方を含んだ概要文であることが望ましいと考える.

そこで, 基本生成手法における 2c ステップに変更を加える．基本生成手法では, 単に高いスコアの断片を選ぶものであったが, 変更手法では, 内容クエリに関連したキーワードと地理クエリに関連したキーワードの双方を含むような断片を優先する．どの断片も双方のキーワードを含んでいない場合は, 内容クエリに関連したキーワードを含む断片と, 地理クエリに関連したキーワードを含む断片を交互に選択していく．交互に選択していく際に, まず最初に選択する断片としては, スコアが高い方を優先する.

3.3 地理クエリを座標として扱う概要文生成法

本節では, 提案システムにおける, 概要文生成方法について述べる．手法では地理クエリおよび文書中の地名表現を座標として扱い, 文書中で地理クエリと強い関係のある地名表現を特定し, 概要文の生成に利用する.

地名抽出時のエラーにより, 文書では本来述べられていない地名が誤って含まれている可能性がある．このようなノイズとなる地名を除去するために, 前処理として, クラスタリングを利用したノイズの除去を行う．詳細は, [2] を参照されたい.

概要文を生成するために有益な地名表現は, 地理クエリと関連性が高いような表現であると考える．このため, 我々は地名表現が示す位置が地理クエリで示される領域に含まれるか, 近接している地名表現を概要文生成に利用する．次節の実験では, 地名表現の代表点が地理クエリの中心点から, 地理クエリで示された範囲の 2 倍以内におさまるような地名表現を用いた.

我々がめざす地理情報検索が, ユーザの居場所のごく周辺の情報を得るためであることを踏まえ, 地理クエリと文書中の地名表現の関連度は, 以下の 2 つの条件に基づき評価する.

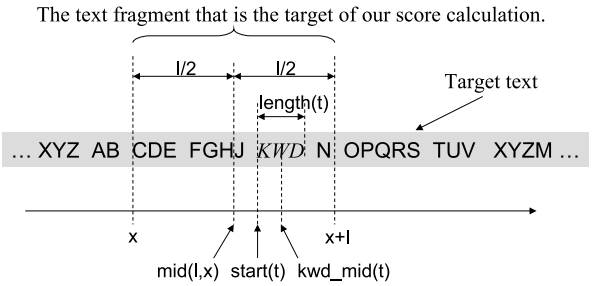
- 地理クエリで示される領域から近いこと
- 特定のエリアについて言及していること

これらの二つはそれぞれ, 「地理クエリが示す場所と地名表現が示す場所の距離」および「地名表現によって示される地理的範囲」によって評価することができる．これらを元にした地名表現 e と地理クエリ Qg の関係の強さ $W(e)$ は以下の式で算出する.

$$W(e)=Closeness(Qg,e)\cdot Specifity(e) \quad (2)$$

$$Closeness(Qg,e)=1/dist(Qg,e) \quad (3)$$

$$Specifity(e)=1/width(e) \quad (4)$$



Case in which "*KWD*" is included in query keywords (*t* = "*KWD*").

図 2　テキスト断片のスコア付けに関する変数

$$dist(Qg, e) = \begin{cases} d_{inner} \\ \quad (Qg_p \text{ is inside } e\text{'s extent}) \\ d_{inner} + d_{edge}(e, Qg) \\ \quad (Qg_p \text{ is outside } e\text{'s extent}) \end{cases} \quad (5)$$

ここで d_{inner} は十分小さな定数であり, 地理クエリの中心点 Qg_p が地名表現 e で示される範囲の中である場合の距離として設定する. また, $d_{edge}(e, Qg)$ は, 地名表現 e が含意する領域の境界線からの最短距離を示す.

上記の地名表現の重み付けを利用し, テキスト断片のスコア S_g を以下の通り定義する.

$$S_g(l,x) = \sum_{e \in E} (W(e) \cdot (\alpha + (1-\alpha) \cdot (1 - \frac{|mid(l,x) - kmid(e)|}{l/2}))) \quad (6)$$

E は対象の文書中に含まれる地名表現のうち, 後述の基準に基いて選択された地名表現 e の集合である. それぞれのテキスト断片の最終的なスコア $S'(l,x)$ は以下の式で算出される.

$$S'(l,x) = S(l,x) + S_g(l,x) \quad (7)$$

上記スコアを用い, 提案システムでは 3.1 節で示した概要文生成手順の 2(b) および 2(d) の $S(l,x)$ をこの $S'(l,x)$ で置き換えた方法によって, 概要文を生成する.

4. 概要文生成評価

検索結果の概要文は, ただし, クエリを受け付けない汎用の要約とは異なり, 検索結果の概要文には, その結果を見てユーザが当該文書を閲覧するかどうかを判断するという明確な目的がある. そのため, 検索結果の概要文の評価は, 概要文そのものを内的 (intrinsic) に評価するのではなく, 概要文がユーザの正しい判断に結びついているか, すなわち概要文がどの程度指示的かに基いた, 外的 (extrinsic) な評価を行う.

地理情報検索では, クエリに内容制約と地理的制約の性質の異なる 2 種類の制約を受け付ける.

評価対象の文書集合として, 「goo ブログ(注1)」から取得したブログ記事を利用した. 取得したブログ記事は, ブログ著者がブログサイトを作成する段階で, ブログサイトの主な記事の内容を「地域情報」もしくは「食べ歩き」とした記事である. このテストコレクションのブログ記事は, 約 19,000 人の著者によって書かれた 296,679 のブログ記事である.

表 1　実験に用いたクエリ

地理クエリ (中心点)	札幌駅, 東京駅, 梅田
地理クエリ (距離)	1km
内容クエリ	カレー, ラーメン, ハンバーグ, ケーキ, コーヒー, 紅茶, 焼肉, うどん

指示性評価のため, 地理情報検索の検索精度の評価のために集収した適合性判定結果を利用した. 検索に用いたのは表 1 に示した 3 地点, 8 キーワードの組合せの計 24 トピックである. この 24 トピックの検索精度の評価のためにプールされた文書数は 1,921 件 (適合: 464 件, 不適合: 1,457 件) であり, これに対して各手法によって生成された概要文について, 以下の基準に基いて人手で判定を行った.

判定基準: 内容クエリで示したキーワードに関連する店舗や組織を, 地理クエリで示した範囲内で探していると仮定した場合に, 当該概要文によって示されている文書をアクセスするかどうか.

閲覧すべきとされた文書が適合文書であるか, 閲覧すべきでないとした文書が不適合文書であるかを判定する. 生成された概要文が十分指示的であれば, 概要文によって, 適切に適合文書を選択できるはずである.

4.1 評価データの性質

我々は, ユーザの居場所の周辺の比較的小さい領域での情報を得るような地理情報検索をめざしている. このような地理情報検索のために, 提案システムでは地名の含意する広さを考慮したランキングおよび概要文生成を行っている. しかし, そもそも検索対象文書がそのような広さを考慮しても意味のないような地理表現しか含んでいない可能性もある. そこで, 文書中の地理表現の出現, またその広さについて調べた. 実験に用いた全 296,679 文書中, 少なくとも 1 つの地名表現が抽出されたのは, 157,713 記事であった. つまり, 実験に用いた記事中 53.2%が潜在的に地理情報検索の検索結果となり得る. 今回実験に用いたブログ記事は著者自身が「地域情報」あるいは「食べ歩き」と

(注1)：http://blog.goo.ne.jp/

表 2　概要文の指示性評価 (適合文書へアクセスできたか)

手法	精度	再現率	F 値
Short			
提案法	0.638	0.642	0.614(**,##)
2 種の制約を考慮	0.671	0.248	0.307(**)
ベースライン	0.338	0.080	0.120
Long			
提案法	0.664	0.624	0.639(**,##)
2 種の制約を考慮	0.637	0.409	0.467(**)
ベースライン	0.653	0.219	0.311

*, # はそれぞれ ベースライン, 2 種の制約を考慮からの有意な改善があったこと示す (Wilcoxon 検定. 危険率は 0.05, 特に**, ## は $p < 0.01$ を示す)

いう内容としたものであるので, このように高い割合になっているものと考えられる. これら 157,713 記事のうち, 62%にあたる 99,067 記事が複数の地名表現を含んでいた. さらに, そのうち 40%にあたる 38,884 記事は地名表現の示す範囲の間に包含関係を持つ地名表現の対を含んでいた. これらの記事について, もし地名の広さによる重み付けを一切行わずに, クエリとの距離やクエリが地名の範囲に含まれるかどうかによってランキングや概要文生成を行なった場合, 広い範囲を示す地名表現に狭い範囲の地名表現が隠蔽されてしまい, 結果的に小さな領域に関して述べた文書と大きな領域に関して述べた文書を峻別できなくなる可能性がある. したがって, 評価に用いたデータは地名の広さを考慮する意味があるデータであると言える.

4.2 比較手法

次に示す手法を比較評価した.

（1） ベースライン: 3.1 節で示した手法. 地理制約は利用されない

（2） 2 種の制約を考慮: 3.2 節で示した手法. 地理制約を文字列として利用.

（3） 提案手法: 3.3 節で示した提案法.

PC 向けと, 携帯端末向けという 2 通りの提示環境を想定し, 最大サイズを 2 種類変えて実験を行った. PC を想定した長い場合は 120 文字, 携帯端末を想定した短い場合は 40 文字とした. この条件は Yahoo!や Google などの検索サービスと同等になるように決定した.

4.3 指示性評価の結果

表 2 に, 概要文を読んだユーザが適合文書にたどり着けるかどうかの評価結果を示す. 表 2 から, 提案手法は F 値において比較手法を大きく上回る評価値を示している. 2 種の制約を考慮した手法の評価値は提案手法と比較して大きく低下しており, 単純に代表的な地名表現を利用するのみでは良い結果を得られないことがわかる.

5. まとめ

本稿では, 地理情報検索において, 特にユーザの居場所の周辺の比較的小さな領域に関して高精度な検索が行える地理情報検索システムを提案した. 地理情報検索は, 従来のキーワード以外にも, 地理的制約を受け付けるため, ランキング方法を変えるだけでなく, 結果の提示においても新しい仕組みが求められる. このため, キーワード文字列のマッチングだけでなく, 地理的制約も考慮した新しい概要文生成方法を提案した. これは我々のめざす小さな領域で高精度な検索に応じた概要文とするため, 地名間の連続的な距離と, 各地名の含意する広さを考慮するものである. 概要文を読んで, ユーザが適切に適合文書を選択できるかどうかについての指示性の評価を行った結果, 提案する概要文生成手法が, 提案手法が比較手法に比べより指示的であることが分かった.

文　献

[1] T. Mandl, P. Carvalho, G. M. di Nunzio, N. Ferro, F. Gey, R. Larson, D. Santos and C. Womser-Hacker: “GeoCLEF 2008: the CLEF 2008 Cross-Language Geographic Information Retrieval Track Overview”, CLEF 2008 (Eds. by F. Borri, A. Nardi and C. Peters) (2008).

[2] 安田, 戸田：“検索位置のごく周辺を対象とした地理情報検索”, 人工知能学会論文誌, **23**, 5, pp. 364–373 (2008).

[3] G. Hobona, P. James and D. Fairbairn: “An evaluation of a multidimensional visual interface for geographic information retrieval”, GIR ’05: Proceedings of the 2005 workshop on Geographic information retrieval, New York, NY, USA, ACM, pp. 5–8 (2005).

[4] B. M. Tomaszewski, C.-C. Pan, P. Mitra and A. M. MacEachren: “Facilitating situation assessment through gir with multi-scale open source web documents”, GIR ’07: Proceedings of the 4th ACM workshop on Geographical information retrieval, New York, NY, USA, ACM, pp. 95–96 (2007).

[5] A. Tombros and M. Sanderson: “Advantages of query biased summaries in information retrieval”, SIGIR, ACM, pp. 2–10 (1998).

[6] A. Turpin, Y. Tsegay, D. Hawking and H. E. Williams: “Fast generation of result snippets in web search”, SIGIR ’07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, New York, NY, USA, ACM, pp. 127–134 (2007).

[7] 平野, 松尾, 菊井：“地理的距離と有名度を用いた地名の曖昧性解消”, 情報処理学会全国大会, pp. 3D–7 (2008).

[8] 竹野, 井上：“分散型高速情報収集/全文検索システム infobee/evangelist”, NTT R&D, **52**, 2, pp. 78–84 (2003).

[9] Y. Chali: “Generic and query-based text summarization using lexical cohesion”, AI ’02: Proceedings of the 15th Conference of the Canadian Society for Computational Studies of Intelligence on Advances in Artificial Intelligence, London, UK, Springer-Verlag, pp. 293–302 (2002).

[10] M. Okumura and H. Mochizuki: “Query-biased summarization based on lexical chaining”, Computational Intelligence, pp. 578–585 (2000).

[11] R. Varadarajan and V. Hristidis: “A system for query-specific document summarization”, CIKM ’06: Proceedings of the 15th ACM international conference on Information and knowledge management, New York, NY, USA, ACM, pp. 622–631 (2006).