

地理範囲とトピックに応じた動的要約生成

An Automatic Summarization Method
that Handles Geographic Information and Topical Information

安田 宣仁 西野 正彬 片岡 良治
Norihiro Yasuda Masaaki Nishino Ryoji Kataoka

日本電信電話株式会社 NTT サイバーソリューション研究所
NTT Cyber Solutions Laboratories, NTT Corporation

We propose a new type of summarization method called *geodigest* that handles geographic information as well as topical information. Geodigest generates an overview of the given geographic area for the given topic from general web search results. Since we do not make special preconditions about documents to be summarized, some sentences in the documents may not have enough geographic-context. To estimate the importance of each sentences or fragments in terms of geodigest, we adopt a pair of geographic information and word as a basic unit of scoring sentences and the score of each pair is defined according to the Z-value, i.e., standard deviation units. To avoid summaries that only describe specific area, we introduce the idea of geographical information gain to the score function. We evaluated the proposed method by ROUGE using hand-made reference data. The proposed method outperforms conventional tf-idf based method.

1. はじめに

近年、高度な携帯端末が普及に合わせて、外出先でその地域に関する情報を探すための各種サービスが普及しつつある。たとえばレストランやホテルの検索などは既に商用サービスとして広く利用されている。こうした、外出先等でその場所に存在すると思われる特定の要素についての情報を知る手段は存在するが、一方で、たとえば初めて訪れた旅先などでは、そもそもその場所がどういった特徴がある場所かを手短かに知りたい場合がある。仮に観光地で観光というトピックに関する情報を知りたいような場合であれば、簡単な検索によって既存の簡潔な説明文が見つかることが多いだろう。しかし、多くの場所についてはそのような説明文は存在しない。また、その他のトピックについて既存の簡潔な説明文があることを期待することもできない。そのような場合、検索を繰り返し、多数の文書を閲覧して所望の情報を得なければならない。

本稿では、トピックと地理範囲が与えられた場合に、その地理範囲の情報を簡潔にまとめた文書を生成するような要約、ジオダイジェストを提案する。ジオダイジェスト生成のポイントは、雑多な文書群から有用な情報抽出すること、そして、網羅的な情報を含んだ要約文を生成することの2点である。前者については、仮にある地域について説明した均質な文書群が入手できるのならば必要のない要件である。しかし、任意の場所についてジオダイジェストを生成するためには、あらゆる場所についての言及を含む大量の文書を用意する必要がある。そのような文書を入手するのは現実的ではないため、web 検索等から得られる雑多な文書集合を要約対象文章集合として想定しなければならない。後者については、通常の自動要約生成問題でも要約文が要約対象の文書群を網羅するものであることは求められているが、それに加え、ジオダイジェスト生成では地理的な網羅性も重要であると考えられる。すなわち、指定された地理的範囲において、局所的な地域の情報だけを含んだ要約となるのではなく、なるべく多くの領域についての情報を含むことである。地理的な網羅性が高い要約を生成することで、ユーザが興味がある地域全体に関して端的に知ることができる。

本稿では文抽出により要約を生成する。上記の2つのポイントを満たす要約を生成するため、本稿では、地理範囲に対する語の出現偏りによるスコア付けと、地理的網羅性のための地名のプレフィックスに基づく冗長度制御を導入する。地理範囲に対する語の出現偏りによるスコア付けでは、場所と語を対として取り扱い、それらの対が他の対に比べてどの程度特徴的に出現しているかを文のスコア付けの基本要素とする。この対を地理情報単位対と呼ぶ。対が、特徴的に出現しているかどうかには、対中の語と地域が、他の地域に比べてどの程度偏っているを用いる。特徴的な対が多数含まれるような文を重要な文として抽出することで、雑多な文書群から地域に関する有用な情報を含む文を取り出すことができると考える。また、地名のプレフィックスに基づく冗長度制御では、正規住所で表現した場合の地名の文字列としての重複にに着目し、他で選択した地理範囲となるべく広く異なる地名を選択し、地理的網羅性を高めた要約を生成する。

上述の地理範囲に対する語の出現偏りを統計的に得るためには、1) 対として認めるに十分な強い結びつきがある出現を、2) 統計として認めるに十分な量集める必要がある。しかし、対としての認定を厳しくすればするほど、対の数は減ってしまう。そこで本稿では強い結びつきの出現として、同一の文内で地名と語が共起した場合とし、統計を得るに十分な量を収集するために、要約対象文書群とは別に準備する大規模な文書群を用いる。

本稿の構成は以下の通りである。まず、2節で提案手法の詳細について述べ、次に、3節で評価方法と評価結果について述べる。4節で関連研究と本稿の関係を示した後、5節でまとめを述べる。

2. 地理範囲とトピックに応じた要約

提案するジオダイジェスト生成手法について説明する。提案する要約手法は事前に、要約のスコア付けの基本単位として利用する地理情報単位対の抽出しておく。要約対象地域と要約対象トピックに応じた文書群を入力とし、入力文書中の語の統計(TF 値)と、抽出した地理情報単位対を用いて要約対象の文書群の文のスコア付けおよび要約の網羅性の評価を行い、最終的な要約を生成する。

2.1 地理情報単位対

地理情報単位対を用いた要約を生成するためには、要約生成に利用するための地理情報単位対を抽出し、抽出した地理情報単位対に重要度を付与する。この手順について述べる。

まず地理情報単位対は、地名と語との文内共起を用いて抽出する。すなわち、ある文中に、地名と何らかの語が同時に出現している場合に、その地名と語とを地理情報単位対とする。たとえば「横浜でケーキを買った」という文があれば、{ 横浜, ケーキ }, { 横浜, 買う }などを地理情報単位対として抽出する。地理情報単位対の抽出方法には、たとえば文書内共起といったより弱いつながりを利用する方法も考えられる。しかし、文書内共起を利用すると本来は抽出すべきでない地名と語の対も誤って抽出してしまう可能性が高くなる。ジオダイジェスト生成に用いるためには精度よく地理情報単位対を抽出する必要があるため、本稿では文内共起を用いた。しかし、文内共起に基づく対の抽出は、抽出される対の数が少なく、十分な統計が得られないという問題がある。そこで本稿では、地名と語の組として認定する対象は、同一文内での共起とするが、統計を取る対象の集合は要約対象文書集合に限定せず、大量の文書から得るという方法を取る。これにより、十分な統計を得るのに使える、精度のよい地理情報単位対を集めることができる。

次に、地理情報単位対の集合と、ジオダイジェストを作成したい対象の地域とが与えられた時に、それぞれの地理情報単位対にスコアを設定する。要約生成は、生成された要約に含まれる概念ユニットのスコアの和が大きくなるように文を選択するプロセスであるから、どのようにスコアを設定するかは重要である。本稿では、ある語のある地域での出現回数が他の地域と比べて多い場合、その語がその地域で特徴的な語であるとして、地理情報単位対に高いスコアを与える。具体的には、語 w の全出現回数を n_w 、地域 l の全出現回数を b_l とし、すべての対の出現の総和が s であるとき、地域に偏りなく語が出現した場合の、地域 l における語 w の期待頻度、すなわち対 $\{l, w\}$ の期待頻度は $n_w \cdot \frac{b_l}{s}$ と計算することができる。これに対する実際の対の出現頻度 $r_{\{l, w\}}$ との差によって対の特徴度合いを決める。二項分布による計算は計算量が大きく困難なため、 $r_{\{l, w\}}$ を標準化し、平均値 $n_w \cdot \frac{b_l}{s}$ 、標準偏差 $\sqrt{n_w \frac{b_l}{s} (1 - \frac{b_l}{s})}$ の正規分布で近似した場合の、 z 値を利用する。すなわち以下の値を用いる。

$$z = \frac{(r_{\{l, w\}} \pm 0.5) - (n_w \cdot \frac{b_l}{s})}{\sqrt{n_w \frac{b_l}{s} (1 - \frac{b_l}{s})}} \quad (1)$$

式中、分子の符号は、対の観測度数が期待度数よりも多い場合に -0.5 、そうでない場合に $+0.5$ である。我々は特徴的に多く出現している対のみを用いたため、上記で算出する z 値が正の対のみを記録し利用する。

2.2 ジオダイジェスト生成

ジオダイジェスト生成は、要約候補である文の集合 D から、スコア文字数制限を満たしつつ、かつスコアが高くなるよう部分集合 $S \subseteq D$ を選択する問題として定式化できる。文集合のスコアは、地理的に見て特徴的な情報をふんだんに含んでおり、かつ、地理的・内容的に冗長でなく、限られた語数の範囲なるべく網羅的であるようにするために、地理情報単位対に基づく文の基本スコア、内容語の網羅性のためのトピック情報ゲイン、および地理的網羅性のための地理情報ゲインによって定義する。以下にそれぞれの計算方法について述べる。

地理情報単位対に基づく文の基本スコア $B(s)$

提案法で利用する基本的なスコアである。各文のスコアを、文中に含まれる地理情報単位対のスコアの和として定義する。その地域において特徴的な単位対が多く含まれるような文ほど文の重要度は高いという考えに基づくスコア付けであり、具体的には、対の地名が要約対象として指定された範囲内であり、しかも対の語が本文中に出現する場合に、その対の z 値と要約対象文書群内の語の頻度 (TF 値) を乗じたものを文のスコアに加算していく。すなわち文 s の基本スコア $B(s)$ を以下の式により定義する。

$$B(s) = \sum_{\text{pair} \in s} ((\text{pair の } z \text{ 値}) \cdot (\text{pair 中の語の TF})) \quad (2)$$

内容語の網羅性のためのトピック情報ゲイン $R(s, S)$

限られた文字数で網羅的な要約を作るためには、冗長な内容を含まないようにする必要がある。そこで、新規に選択しようとする文 s 中の語集合 W_s のうち、既に選択されている文集合中の語集合 S に含まれていない数を用いて評価する。すなわち文 s のトピック情報ゲイン $R(s, S)$ を以下の式により定義する。

$$R(s, S) = |W_s \setminus S| \quad (3)$$

地理的網羅性のための地理情報ゲイン $G(s, S)$

1 節で述べたように、ジオダイジェスト生成においては内容的な網羅性に加えて地理的な網羅性も考慮した要約を作成する必要がある。地理的に網羅的な要約とは、つまり地理的な冗長性が小さい要約のことをいう。たとえば、神奈川県に関するジオダイジェストを生成することを考えた場合、神奈川県横浜市に関する内容しか含まない要約は、地理的な冗長度が高い要約である。また、異なる 2 つの地名であれば、より広い範囲を示す語の方が地理的観点において網羅的な要約を生成するには有効であると考えられる。我々は、(日本国内の) 正規住所は前方ほど上位の階層であり、共通の階層についてはプレフィックスを共有するとに着目した。新たな地名は、既に選択されている地名となるべく共有する部分が少なく、かつ、新たに含まれる地名はなるべく短いものを優先することで、重複が少なく、かつ全体を網羅しやすい選択が可能となると考えられる。そこで、新規に選択しようとする文中の地名の正規住所を t 、既に選択された文集合を S 、 S と関連付けられた地名の正規住所集合を L 、地名 x の長さを、 len_x 、ふたつの地名の前方からの共通文字数を len_{com} とするとき、以下の式で文 s の地理情報ゲイン $G(s, S)$ を定義する。

$$G(s, S) = \min_{t \in L} \frac{(\text{len}_t - \text{len}_{com}) + (\text{len}_t - \text{len}_{com})}{\text{len}_t} \quad (4)$$

文スコアの算出と要約生成

このようにして算出した上記 3 つの値を用いて、文 s のスコア $\text{score}(s)$ を以下のように定義する。 λ_1 および λ_2 はそれぞれトピック情報ゲイン、地理情報ゲインの重みパラメータである。

$$\text{score}(s) = B(s) + \lambda_1 R(s, S) + \lambda_2 G(s, S) \quad (5)$$

Algorithm 1 に N 文からなる文集合 D から長さ制限 L_{max} で要約を生成するアルゴリズムを示す。基本的には、上記のスコアが最大となるような文を文字数制限に達するまで、貪欲法により順次選択することによって行う。

Table 2: 各手法での ROUGE-1 スコア (3 つの参照要約での値の平均)

手法	平均	近畿	広島県	群馬県	名古屋市	松山市	梅田
トピック: 「歴史」							
提案法	0.211	0.223	0.178	0.265	0.248	0.197	0.152
TF-IDF	0.202	0.209	0.238	0.306	0.131	0.237	0.088
トピック: 「食べ物」							
提案法	0.250	0.225	0.291	0.250	0.248	0.182	0.302
TF-IDF	0.197	0.206	0.179	0.123	0.189	0.195	0.270

Algorithm 1 要約生成手順

Input: Set of sentences $D = \{s_1, \dots, s_N\}$, L_{max}

Initialize: $C \leftarrow \{1, \dots, N\}$, $S \leftarrow \emptyset$, $\ell \leftarrow 0$

```

while  $C \neq \emptyset$  do
   $maxscore \leftarrow 0$ 
   $i_{max} \leftarrow -1$ 
  for  $i \in C$  do
     $score \leftarrow B(s_i) + \lambda_1 R(s_i, S) + \lambda_2 G(s_i, S)$ 
    if  $score > maxscore$  then
       $maxscore \leftarrow score$ 
       $i_{max} \leftarrow i$ 
    end if
  end for
  if  $i_{max} < 0$  then
    break
  end if
  if  $\ell + \text{length}(s_{i_{max}}) \leq L_{max}$  then
     $S \leftarrow S \cup \{s_{i_{max}}\}$ 
     $\ell \leftarrow \ell + \text{length}(s_{i_{max}})$ 
  end if
   $C \leftarrow C \setminus i_{max}$ 
end while
return  $S$ 

```

Table 1: 実験に使用したトピックと地域

トピック	「歴史」, 「食べ物」
地域	「近畿」, 「広島県」, 「群馬県」, 「名古屋市」, 「松山市」, 「梅田」

3. 評価

提案手法の有効性を評価するため、人手で作成した参照要約を用いた ROUGE[Lin 04] スコアの比較を行った。

3.1 実験条件

地理情報単位対の抽出対象としての大規模文書群として、ブログ約 1 億記事を用いた。地名の認定と正規の住所の割り当てには、平野ら [hir 08] の手法を用いた。対の単位としては、地名は正規住所とし、語はバイグラムとした。対の認定は同一文内での共起のみとし、下位階層の地名での共起回数も上位階層での地名の共起回数に加えた。つまり、たとえば「神奈川県横須賀市光の丘」という地名と「通研-で」というバイグラムが 1 度出現した場合、{ 神奈川県, 通研-で } や { 神奈川県横須賀市, 通研-で } といった対の数も 1 度ずつ数える。なお、地理情報単位対の抽出対象はブログである必然性はなく、文区切りや形態素解析が一般の文書に比べて難しいという点ではむしろ一般の文書の方が好ましいと考えられる。しかし、ブログは RSS を利用して大量の文書を容易に収集することができるため、本実験ではブログを用いた。

評価に用いた課題は、表 1 に示す通り、2 種類のトピックと 6 つの地域である。トピックは出張や旅行先といった外出先で興味を持つような典型的な対象として選択した。また、地域の選択基準は、広い地域と狭い地域、都会と田舎、単一の地名と複数の住所からなる通称がいずれも含まれるように選択した。

要約対象文書群は、トピックと地名を検索キーワードとした一般の web 検索サービス (Google) による検索結果上位 100 件を用いた。web から取得した文書には、本文以外の情報が含まれているため、ExtractContent^{*1} を用いて、本文と思われる部分のみを取り出して処理対象とした。

また、web ページからの雑多な文であるため、文区切りを自動で正確に行うことは困難である。このため、人間が見れば複数の文から構成されているような文が単一の文として認定されてしまったり、逆に、単一の単語のみが文として認定されてしまう場合がある。このため、実験では、長さが上位下位それぞれ 5% 未満の文は外れ値として事前に除去した。

人手による参照要約は、各地域の要約対象記事各 100 記事を用いて、ひとつの課題につき異なる 3 名で 3 種類の要約を作成した。総計 15 名で作成し、背景知識をなるべく使わずに与えられた文書群のみから参照要約を作るため、それぞれ担当する地域にはあまりなじみがなく、居住実績がないことを条件とした。参照要約の長さ制限は 500 文字 $\pm 5\%$ とした。参照要約は文抽出による要約ではなく、自由記述であるが、固有名詞については要約対象文書中に出現する表記と揃えるようにした。また、要約にあたって、なるべく地域の偏りなく長さ制限の中で網羅的であるようにという教示を行った。

ベースライン手法は、文のスコアとして TF-IDF を使い、内容語の冗長性の抑制のために提案法同様の貪欲法による抑制を行う手法とした。IDF の算出には、提案法における地理情報単位対の抽出対象同様、ブログ約 1 億記事を用いた。

提案法で用いる 3 つのスコア $B(s)$, $R(s, S)$, $G(s, S)$ はいずれも最大のスコアが 1 となるように正規化した上で、パラメータである式 5 中の、 λ_1 , λ_2 はいずれも 1 とした。同様に、ベースラインについても、TF-IDF スコア、 $R(s, S)$ はいずれも最大のスコアが 1 となるように正規化し、2 つの重みは等しいとした。

3.2 評価結果

表 2 に提案法とベースラインでの ROUGE-1 スコアの平均を示す。今回作成した参照要約は、ひとつの課題 (タスクと地域) について 3 つの参照要約が存在するので、それぞれとの ROUGE-1 スコアの平均を示した。表より両トピックについて平均スコアにおいて提案法がベースラインを上回っていることが分かる。また、平均値だけでなく、トピック「歴史」については 6 地域中 5 地域について上回っており、提案法は与えられたトピックについての地域の要約において、より人手による

*1 <http://labs.cybozu.co.jp/blog/nakatani/2007/09/web.1.html>

参照要約	松山市は、平成 12 年 4 月 1 日に中核市となり、平成 17 年の 1 月 1 日に行われた北条市と中島町との合併後は、人口 514,784 人（平成 18 年 3 月 1 日現在）となっている。道後湯之町は、3000 年の歴史がある道後温泉がある。この温泉は別名・坊っちゃん湯と呼ばれ親しまれる街のシンボリック存在で、1994 年に国の重要文化財（文化施設）として指定、2009 年には経済産業省の「近代化産業遺産」にも認定されている。太山寺町は、天平 11 年、聖武天皇の勅願によって行基が十一面観音像と四天王像を安置したと伝えられている太山寺がある。和気町は天平年間、聖武天皇の勅願によって行基が開いた古刹であり、寛永 10 年、須賀重久が荒廃していた堂宇を再建した円明寺がある。保免は 1400 年以上の歴史を持ち、「落日を呼び戻した」とされる伝説を持つ日招八幡神社がある。南江戸は享保十八年、松山藩の危機を救った純忠至誠の士「山内与右衛門久元」を祀った神社がある。山越は 900 年以上の歴史を持つ還熊八幡神社がある。道後湯之町は伊予・松山城とともによく知られた平山城で建武年間に河野通盛によって築かれた湯築城がある。
提案法	旅館内に露天風呂があるとか、家族風呂があるとか温泉と言っても様々なものがあるのです。ここからしばらく歩いた先にある雲巖寺には武蔵が『五輪の書』を書いたとされる霊巖洞があります。桜の木は枯れたが孝子吉平の話として長く伝えられ、龍穩寺裏山の桜谷、吉平邸跡といわれるところにある早咲き の桜と天徳寺にある桜がこの名で呼ばれている。この時代の遺跡を代表するものに釜ノ口遺跡があり、更に木器や高床式建物の一部である鼠返し（ねずみかえし）を出土した朝美遺跡がある。若手県中尊寺及び新潟県弥彦神社にも同人奉納の銅板納札があるが江戸初期の民間信仰を知るうえで貴重な文化遺産である。鱒施餓鬼（どじょうせがき）のこと税金などに苦しんでおりましたが、皆で坂の上の雲にあるような国と戦うため、この国を守るためには、一人一人が冬期は、九州の日本海側にある雪雲が関門海峡を抜け、松山上空まで雪を降らせることがある。この奥に霊巖洞があります。霊巖洞は雲巖寺の奥之院、五百羅漢を右手に見ながら歩いた一番奥にあります。霊巖洞は宮本武蔵が『五輪の書』を書いたとされる場所です。眼の後に 7 個の鱧孔エライナがある。
TF-IDF	愛媛県美術館（愛媛県松山市）[美術] 鹿島博物館（愛媛県松山市）[自然] 既に内閣総理大臣認定を素欧詠橋 樑梁渦回目は、愛媛県松山市の『坂の上の雲』のまち再生計画を御紹介します。（松山市の課題）松山市北条ふるさと館 歴史民俗展示室・美術展示室（ジャンル：歴史・民俗 / 愛媛県松山市）愛媛県松山市堀之内 089-934-1534 愛媛県歴史文化博物館 桜の木は枯れたが孝子吉平の話として長く伝えられ、龍穩寺裏山の桜谷、吉平邸跡といわれるところにある早咲き の桜と天徳寺にある桜がこの名で呼ばれている。 愛媛県松山市の松山城の北東、マイカーで市内より 15 分、伊予鉄路線バスで、市駅より 30 分の位置にあります。松山市北条ふるさと館 歴史民俗展示室・美術展示室（愛媛県松山市）[歴史・民俗] 協力：Mapion: 松山市水道資料館（愛媛県松山市）[理工] 松山市考古館（愛媛県松山市）[歴史・民俗] 迷路絵本作家で歴史考証イラスト画家の香川元太郎さん（愛媛県松山市出身）の迷路作品（原画・巨大迷路パネル等）をはじめ、愛媛の歴史イラスト等を展示・紹介する。愛媛県松山市の 桜の名所

Figure 1: 要約出力と参照要約の例（トピック:「歴史」、地域「松山市」、長さ制限 500 文字）

参照要約に近い要約を生成していることが分かる。ただし、提案法が上回ると言っても ROUGE-1 スコアが高々 0.25 程度と低い。この要因として、本実験での設定の困難さが挙げられる。まず、本実験では要約対象文書は何の制約もない一般の web 検索結果であり、内容も品質も雑多である。また、今回要約対象文書の数 100 (平均文数約 5,000) と従来複数文書要約で用いられてきたデータセットに比べ多い。これらの困難さがスコアが低い要因のひとつであると考えられる。また、提案法によるスコアは IDF と排他的ではなく、組み合わせることでより精度の高い要約を生成できる可能性がある。

図 1 に参照要約と各手法での要約出力の例を示す。一見したところ優劣をつけ辛いが、提案法の出力には、この地域（松山市）に特徴的な語が多く含まれていることが分かり、我々の目的である、知らない場所についてそこがどのような場所かを手短かに知るといった目的に近い結果であると見ることができる。

4. 関連研究

地理的な観点での要約としては、Chakraborty ら [Chakraborty 11] によって、気候や農業に関する場所特有の要約方法が提案されている。しかし、本稿でめざすような地理的になるだけ広い範囲についての言及を含む方策は取られていないため、地理的に冗長な要約が生成され得る。

冗長性の排除は従来より要約においての主要な課題のひとつであり、逐次的に選択する際に冗長性を考慮する方法 [Goldstein 00] のみでなく、近年では、組み合わせ最適化問題として定式化されることが多い [McDonald 07, Filatova 04, Takamura 09]。本稿では、地域の観点からの冗長性、および、内容の観点からの冗長性を制御するためには貪欲法を用いたが、地理的条件も含めた組み合わせ最適化として定式化することも可能であると考えられる。

5. おわりに

地理範囲とトピックの双方を考慮した新しい要約「ジオダイジジェスト」を提案した。提案法は各文の地理的特徴をスコアに勘案するために、また、地理的な網羅性を高めるために、地理情報ゲインによる地理的冗長性の制御を行う。性質の異なる 6 地域に関して「歴史」と「食べ物」というトピックについて、人手で作成した参照要約を用いて、ROUGE-1 スコアによる評

価を行い、提案法によってスコア付けした要約が、ベースラインよりも上回ることを確認した。

References

- [Chakraborty 11] Chakraborty, S. and Subramanian, L.: Location specific summarization of climatic and agricultural trends, in *Proceedings of the 20th international conference companion on World wide web*, pp. 463–472, New York, NY, USA (2011), ACM
- [Filatova 04] Filatova, E. and Hatzivassiloglou, V.: A Formal Model for Information Selection in Multi-Sentence Sentence Extraction, in *Proceedings of the 20th COLING* (2004)
- [Goldstein 00] Goldstein, J., Mittal, V., Carbonell, J., and Kantrowitz, M.: Multi-Document Summarization by Sentence Extraction, in *Proceedings of ANLP/NAACL Workshop on Automatic Summarization* (2000)
- [hir 08] 平野 徹, 松尾 義博, 菊井 玄一郎: 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)
- [Lin 04] Lin, C.-Y.: ROUGE: A Package for Automatic Evaluation of Summaries, in *Proc. of Workshop on Text Summarization Branches Out*, pp. 74–81 (2004)
- [McDonald 07] McDonald, R.: A Study of Global Inference Algorithm in Multi-document Summarization, in *Proceedings of the 29th ECIR* (2007)
- [Takamura 09] Takamura, H. and Okumura, M.: Text Summarization Model based on Maximum Coverage Problem and its Variant, in *Proceedings of the 14th EACL* (2009)