

適合性フィードバックに基づく検索意図を反映させた文書ランキング

畠中 翔太[†] 島田 諭^{††} 三浦 孝夫[†]

[†] 法政大学大学院 工学研究科 〒184-8584 東京都小金井市梶野町 3-7-2

^{††} 法政大学マイクロ・ナノテクノロジー研究センター 〒184-0003 小金井市緑町 3 - 1 1 - 1 5

E-mail: [†]syokiti009@yahoo.co.jp, ^{††}satoshi.shimada.57@hosei.ac.jp, ^{†††}miurat@k.hosei.ac.jp

あらまし 多くの検索エンジンは、あらかじめ決められたランキングアルゴリズムによって、Web ページや文書の集合から抽出した特徴（重要度、トピック）やクエリなどから抽出したユーザの検索意図などを用いてユーザの検索意図に近い文書をランキングする。しかし、ユーザは検索システムに対して数語程度のキーワード入力しか行なわないため、検索意図を抽出するのは難しい。また、文書集合の特徴が必ずしもユーザの重要度やトピックの定義に一致しているとは限らない。本研究では、ユーザに与えた検索結果に対するユーザの意図をユーザには少ない負担でシステムにフィードバックさせることで、検索結果のリランキングを行い、検索結果の精度を向上させる。

キーワード 適合性フィードバック, ランキング, 検索意図

Shota HATAKENAKA[†], Satoshi SHIMADA^{††}, and Takao MIURA[†]

[†] Dept. of Elect. and Elect. Engineering, HOSEI University

3-7-2, KajinoCho, Koganei, Tokyo, 184-8584 Japan

^{††} Research Center for Micro-nano Technology, HOSEI University

3-7-2, KajinoCho, Koganei, Tokyo, 184-8584 Japan

E-mail: [†]syokiti009@yahoo.co.jp, ^{††}satoshi.shimada.57@hosei.ac.jp, ^{†††}miurat@k.hosei.ac.jp

Abstract

Key words

1. 前書き

近年、インターネット上では Blog や twitter など、多様で大量の情報が蓄積されている。しかし、ユーザは検索対象の文書集合について知らないため、どのようなクエリを与えれば、適切な検索結果が返されるかわからないため、ほとんどのユーザは短いクエリや曖昧なクエリを与えてしまう。ユーザが必要とする情報を効率よく入手するためには、これまで主流だった情報検索だけでは必ずしも対応しきれない状況となってきた。

これまでの情報検索は、基本的には入力されたクエリと文書とのマッチングにより、候補文書をユーザに返すものである。クエリと文書のマッチングにおいては、単語の出現頻度などの情報が中心的に用いられてきた。しかし、ユーザの検索意図は多様であり、必ずしも出現頻度に基づく重み付けだけで対応できるとは限らない。これは、検索キーワードが同音異義語にある場合に起こる。例えば、検索者が「マウス」で検索したときにコンピュータのポインティングデバイスのマウスについての文書を検索したいが、ネズミや口の「マウス」の出現頻度が高いために、検索結果の上位に検索意図とは関係ない文書が上位にランキングされてしまう。ユーザや利用環境の多様化により、

ユーザが常に適切なクエリを入力できるということも期待できなくなっている。

このような問題を解決するための手法は多く提案されている。検索対象の文書集合において効果的な検索が可能である有用なクエリを提示、追加する「クエリ拡張」がある。クエリ拡張は、情報検索を実用化するために重要な技術であり、様々な手法が提案されている。例えば、文書クラスタおよび単語クラスタからトピックを抽出し関連語で拡張する手法、相対的にクリック数が多いもの（人気があるもの）に誘導による拡張手法、ユーザの閲覧履歴をもとに、ユーザの嗜好（検索履歴）に合わせてパーソナライズな拡張手法などがあるが、予め決められた共通のテーマやトピックは存在しないために正確なクラスタリングは難しく、履歴情報の不足による精度の低下などの問題がある。

また、ユーザが与えたクエリで得られた検索結果からどの文書が適合・不適合をシステムに学習させることで、ユーザにとって有用なクエリに改善する「適合性フィードバック」がある。適合性フィードバックは、3つのタイプに分けることができる。明示的フィードバックは、あるクエリによって検索された初期結果から適合文書と不適合文書をユーザに選択してもらうことで、クエリを修正する。暗示的フィードバックは、検索

結果の文書をユーザがどの文書を閲覧したか、閲覧時間、ブラウジング、スクローリングなどの行動から適合情報を抽出する。疑似フィードバックは、あるクエリによって検索された文書ランキングの上位数件を適合文書であると仮定することで、クエリを修正する。

また、確率モデルやクラスタリングを用いた研究もあり、文書が有する文脈をトピックモデルを用いて推定し、同時にクエリの意図を推定して適合させる研究やクラスタを作成し、そのクラスタを用いて適合性フィードバックを行う研究などもある。

しかし、適合文書や不適合文書の選択による負担、システムの重要度とユーザの重要度が不一致などで検索結果を改善するにはユーザに大きく負担がかかってしまう。また、初期検索の性能に大きく依存してしまうという問題や予め決められたトピックは存在しないために正確なクラスタリングは難しい。

本研究では、検索結果からユーザの意図を抽出しやすいように、クエリに関連する複数のトピックを含む文書を上位にランキングさせる。その文書を上位にランキングさせることで、ユーザの検索意図にあった情報を早い段階で入手させ、またユーザがフィードバックする負担を減らし、検索結果のリランキングを行うことで検索結果の精度を向上させる。また、提案手法の有効性を示すため、Rocchio アルゴリズムとの比較を行う。

2. Rocchio アルゴリズム

ベクトル空間モデルとは、大量の情報の中から利用者が与えた質問に対して利用者が必要と思われる文書の集合を提示するデータ表現方法である。

データ記述では文書を単語の多重集合で表現する。単語の並びを無視し、文書を多次元ベクトルとして表す。このモデルで重要な語を扱うには十分である。位置情報を失うため精密な表現ではないが、意味内容を表現しない語や記号などを無視できる。文書 d をベクトル $d = \langle v_1, \dots, v_n \rangle$ で表すとき $i = 1, \dots, n$ は予め定まった語 w_i を表し、 v_i はその語に与える重みを意味する。重みとしては、2 値や出現頻度、また TF*IDF 値が用いられることが多い。

またベクトル空間モデルでは、情報検索操作を自然に表現することができる。実際、ベクトルの各要素は語の重みを表すため、当該部の要素を対応させればよい。データベース検索と異なり、情報検索では、完全一致解よりも、質問に多く関連する文書を探索することが要求される。質問に類似する度合いを数値で表現し、この度合い順に文書をランキングして提示する。

例えば、与えられた質問をベクトル化した質問ベクトル q と文書の単語をベクトル化した多次元の文書ベクトル d_i のなす角度により類似度を計算する余弦類似度がある。余弦類似度を $\cos(d_i, q)$ とすると、 $\cos(d_i, q)$ は次式で表せる。

$$\cos(d_i, q) = \frac{d_i \cdot q}{|d_i| |q|}$$

質問ベクトル q と全ての文書ベクトル $d_1, d_2 \dots d_n$ の余弦類似度を計算し、余弦類似度の大きいほど質問に似た文書であることがわかる。

Rocchio アルゴリズム [1] は、ベクトル空間モデルを用いて適合性フィードバックを行い、クエリベクトルを更新することで検索結果の適合率を上げる手法である。例えば、ユーザが与えたクエリベクトル q に対する文書ランキングの中に含まれる適合文書集合を D_r 、不適合文書集合 D_n とする。ユーザは下の式を用いて、ユーザが適合と判断した文書に含まれる語の重みを出現回数や TF*IDF 法などにより大きくし、逆に不適合文書に含まれる語の重みは小さくすることで、クエリベクトル q を q' に修正する。

$$q' = q + \frac{1}{|D_R|} \sum_{d_i \in D_R} d_i - \frac{1}{|D_N|} \sum_{d_j \in D_N} d_j$$

上式で、 $\frac{1}{|D_R|}$ 、 $\frac{1}{|D_N|}$ は重み付け係数であり、 $|D_R|$ および $|D_N|$ は、それぞれの文書集合に含まれる文書数である。

3. 提案手法

3.1 単語の関連度

本研究では、検索結果からユーザの意図を抽出しやすいように、クエリに関連する複数のトピックを含む文書を上位にランキングさせるため、クエリと各単語との関連度を抽出する。この関連度が高い単語を多く含む文書は、クエリに関連する複数のトピックを含む文書となる。

各単語の 1 日の文書頻度をヒストグラムで表し、各ヒストグラム間の類似度を計算する Bhattacharyya 係数 [2] を用いることで、各単語間の関連度を抽出する。Bhattacharyya 係数とは、二つの正規化したヒストグラム間の類似度の計算などに用いられている。式は以下で表される。

$$Bha = \sum_{u=1}^m \sqrt{P_u Q_u} (0 \leq Bha \leq 1)$$

ここで、 p, q は比較対象となる正規化された ($\sum_{u=1}^m P_u = \sum_{u=1}^m Q_u = 1$) ヒストグラムを用いて算出する。ある事象が発生したとき、電子文書、blog、twitter などは、文書ストリームを時系列に沿って観測すると、ある期間において、ある単語を含む文書の時間軸方向の密度が高くなるような状態が発生する。クエリに関連した単語はクエリと類似した時間軸方向の密度が高くなるので、クエリに対して Bhattacharyya 係数が高い語は関連語にある。

しかし、クエリと共起性の高い単語は、ヒストグラムが類似しているため、Bhattacharyya 係数では関連語になってしまうので、共起度の逆数をかけることで、クエリ q と関連のあるトピックに含まれる単語との関連度は高くなる。クエリ q と単語 i の関連度 BC_{iq} は、次式で表される。

$$BC_{iq} = Bha_{iq} * \log\left(\frac{1}{CO_{iq}}\right)$$

ここで、 Bha_{iq} はクエリ q と単語 i の Bhattacharyya 係数、 CO_{iq} はクエリ q と単語 i の共起度である。クエリが 2 個以上のときは、 Bha と CO はそれぞれ掛け合わせて用いる。また、縦軸を Bha 、横軸を単語にして Bha が左側から高い順に並べ

たとき、曲線の傾きが急になる単語の Bha 以上の単語のみをクエリとの関連語として用いる。 CO_{iq} は次式で表される。

$$CO_{iq} = \frac{c}{a+b-c}$$

ここで、 a は単語 i を含む文書数、 b はクエリ q を含む文書数、 c は単語 i とクエリ q を含む文書数である。

3.2 フィードバック

本研究ではユーザがクエリ q を与えたとき、クエリ q と各単語の関連度 BC を用いて全文書にランク値 dr を与え、検索結果の上位 N 件を用いて適合性フィードバックを行う。

ユーザが閲覧した文書を適合文書、閲覧しなかった文書を不適合文書と見なし、単語の関連度を修正していく。そして、新しい単語の関連度を用いてリランキングを行う。単語の関連度の更新式は、次式で表される。

$$BC_{iq,n+1} = BC_{iq,n} \left(1 + \frac{w_{i,D^+}}{|N^+|} - \frac{w_{i,D^-}}{|N^-|} \right)$$

$BC_{i,n}$ は、 n 回目のクエリ q と単語 i の関連度である。 N^- 、 N^+ は n 回目の検索での適合文書数、不適合文書数であり、 w_{i,D^+} 、 w_{i,D^-} は適合文書集合での単語 i の出現頻度、不適合文書集合での単語 i の出現頻度を示している。 $n+1$ 回目の文書 j のランク値 $dr_{j,n+1}$ は、次式で表される。

$$dr_{j,n+1} = \frac{\sum_{n=1}^N tf_{i,j} * BC_{iq,n+1}}{N}$$

$tf_{i,j}$ は文書 j での単語 i の出現頻度、 N は文書 j の単語数である。また、文書 j の共起度 CO_j が閾値以下の文書はランキングから除外する。式は以下で表される。

$$CO_j = \frac{\sum_{n=1}^N tf_{i,j} * CO_{iq}}{N}$$

3.3 手順

本研究の適合性フィードバックの一連の流れを説明する。

(1) ユーザがクエリをシステムに与えるシステムはクエリ q と単語 i の関連度 BC_{iq} を用いて、クエリに関連する複数のトピックを含む文書を上位にランキングさせる。

(2) ユーザは検索結果の上位 N 件から適合性フィードバックを行うクエリに関連する複数のトピックを含む文書含む検索結果からフィードバックを行うため、検索意図の曖昧なクエリからも検索意図を抽出できる。システムはユーザのフィードバックから $BC_{i,n}$ を算出し、文書集合の特徴をユーザの好みに訂正する。

(3) ユーザにリランキングした検索結果を与える

この一連の流れからユーザの検索意図にあった情報を早い段階で入手させ、ユーザの負担を減らし、検索結果のリランキングを行うことで検索結果の精度を向上させる。

4. 実験

4.1 実験方法

実験では、新聞記事集合に対して提案手法を行う。

毎日新聞記事データ集 2010 年版から、1 面から 3 面に掲載

された記事 1 年分、7,999 記事を用いて実験する。記事のタイトルおよび本文を形態素解析し、名詞、動詞、未知語を使用する。名詞については、以下に示すルールで複合語として扱う。

- ・連続する名詞は結合する
- ・接尾語に後続する語は結合しない
- ・数字は“ * ”に置き換える

これらのルールにより、人名はフルネームで 1 語とし、数字を含む語は数字の桁数にのみ着目する。クエリ拡張において、極端な低頻度語および高頻度語は有用ではないと考えられることから、抽出された単語の文書頻度が 20 以上 1,000 未満となる 6,788 語を使用する。

4.2 評価方法

評価には、平均適合率を用いて Rocchio アルゴリズムとの比較を行う。3 つのパラメータ α 、 β 、 γ は、一般的な $\alpha=1.0$ 、 $\beta=0.8$ 、 $\gamma=0.1$ を使用する。初期検索の検索結果の上位 7 件に対して、適合不適合の判定を行う。適合不適合の判定は、ユーザの検索意図に合っている単語が、タイトルや第 1 段落に含まれていれば適合文書とする。例えば、クエリ“中国漁船衝突事件”のみ与えたユーザが“ビデオ流出事件について知りたい”という検索意図だったときは、“ビデオ流出”、“ビデオ映像 流出”、“ビデオ映像 流出”などが含まれていれば適合文書になる。

4.3 実験結果

表 1 は Rocchio アルゴリズムと提案手法で、3 回フィードバックを行った時の top10 での適合文書数を表している。

また、クエリの横のカッコ内はユーザの検索意図、R) は Rocchio アルゴリズムの結果を表している、3 回目のフィードバック時では、5 クエリ中 3 クエリで提案手法が Rocchio アルゴリズムより高い適合率を算出している。

4.4 考察

実験結果から、提案手法について考察を行う。

本研究では、検索結果からユーザの意図を抽出しやすいように、クエリに関連する複数のトピックを含む文書が上位にランキングさせたが、表 3 からわかるように、記事 7815 はクエリ“中国漁船衝突事件”に関連する複数のトピックを含む文書が上位にランキングされている。表 2 の Rocchio アルゴリズムの初期検索結果では、7 文書中 5 文書が“ビデオ流出”についての文書のため、ユーザが“ビデオ流出”以外のトピックについて検索する場合は、時間がかかってしまう。

実際に、表 1 でのクエリ“中国漁船衝突事件(中国人船長)”では、適合文書が 1 つしかないのと不適合文書が“ビデオ流出”についての文書が多いために、“ビデオ流出”に不適合情報はシステム側に大きく反映されたが、その他の不適合情報が反映されないために、ユーザの検索意図がシステムに反映されるのに時間がかかってしまっている。

表 4 と表 5 からわかるように Rocchio アルゴリズムは余弦ベクトルで文書を検索するために、文書の単語数が少ない場合やクエリの出現頻度が文書内で高いなどが原因でクエリ“民主大敗”を与えても、“民主”のみや“大敗”のみを含む文書でもランキングの上位に順位付けしてしまう。そのため、適合文書

が上位にないためにフィードバックが出来ない。

提案手法は、クエリと関連度の高い単語を用いて文書ランキングを行うので、トピック “参院選挙で民主党が大敗” に含まれる単語はクエリと高い関連があるので、このトピックに関する文書が上位にランキング出来ている、また、仮にこのトピックに関連する文書がクエリの単語を含んでいなくても、時系列に単語間の関連性を算出しているため、表6のようにフィードバックを繰り返すことでトピックに関連する文書はランキング上位に上位にランクされる。

5. 結 論

本研究では、検索結果からユーザの意図を抽出しやすいように、クエリに関連する複数のトピックを含む文書を上位にランキングさせることで、ユーザの検索意図にあった情報を早い段階で入手させ、またユーザがフィードバックする負担を減らし、検索結果のリランキングを行うことで検索結果の精度を向上させた。また、時系列に単語間の関連性も求めているため、クエリを含んでいなくても検索意図に合った文書を上位にランキングできる。

文 献

- [1] Rocchio, J.J. Relevance feedback in information retrieval. *The SMART Retrieval Systems*, pp313-323, Prentice-Hall, 1971.
- [2] Qingshan LIU and Dimitris N. METAXAS. Unifying Subspace and Distance Metric Learning with Bhattacharyya Coefficient for Image Classification. *Lecture Notes in Computer Science*, 2009, Volume 5416/2009, 254-267.
- [3] J. Kleinberg. Bursty and hierarchical structure in streams. *In Proc. 8th SIGKDD*, pp. 91-101, 2002.
- [4] Shota HATAKENAKA and Satoshi SHIMADA and Takao MIURA. Ranking Documents with Query-Topic Sensitivity. *International Workshop on Web Information Retrieval Support Systems in IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, 2012.

表 1 提案手法と Rocchio の適合文書数

クエリ (検索意図)	FB[0]	FB[1]	FB[2]	FB[3]
R) 中国漁船衝突事件 (ビデオ流出)	8	9	10	10
R) 中国漁船衝突事件 (仙谷由人官房長官)	2	7	7	7
R) 中国漁船衝突事件 (中国人船長)	1	3	1	1
R) 民主 大敗 (参院選挙で民主党が大敗)	0	0	0	0
R) 郵便不正事件 (虚偽有印公文書作成・同行使罪に問われた事件)	3	9	9	10
中国漁船衝突事件 (ビデオ流出)	5	8	8	10
中国漁船衝突事件 (仙谷由人官房長官)	1	5	6	10
中国漁船衝突事件 (中国人船長)	1	1	1	2
民主 大敗 (参院選挙で民主党が大敗)	2	3	4	4
郵便不正事件 (虚偽有印公文書作成・同行使罪に問われた事件)	5	10	10	10

表 2 Rocchio アルゴリズムのクエリ”中国漁船衝突事件”の初期検索結果

記事	タイトル	本文
6776	中国漁船・尖閣領海内接触：ビデオ流出 菅首相、原因究明指示	菅直人首相は 5 日午前の閣僚懇談会で、中国漁船衝突事件の映像の流出問題について、「しっかり調査して、原因究明しなければならない」と情報管理態勢の点検を指示した。
6812	中国漁船・尖閣領海内接触：ビデオ流出 鳩山前首相「クーデターだ」	民主党の鳩山由紀夫前首相は 6 日、中国漁船衝突事件の映像流出問題について「情報流出によるクーデターのようなことを政府内の人間が行うのは政権にとって大変厳しい話だ」と述べた。
6816	中国漁船・尖閣領海内接触：ビデオ流出 鳩山前首相「クーデターだ」 【大阪】	民主党の鳩山由紀夫前首相は 6 日、佐賀市で講演し、中国漁船衝突事件の映像流出問題について「情報流出によるクーデターのようなことを政府内の人間が行うのは政権にとって大変厳しい話だ」と述べた。その上で「国全体で信頼をつくり上げていくことが大切だ」と強調した。
5728	中国漁船・尖閣領海内接触：SMAP 上海公演のチケット販売停止	【上海・鈴木玲子】中国漁船衝突事件で日中関係が悪化する中、上海で 10 月 9、10 の両日に開催される予定の人気グループ「SMAP」のコンサートのチケット販売が停止されていたことが 20 日分かった。現地代理店は「主催者側からの通知により 18 日から停止した」と説明しているものの、公演自体が中止される可能性も認めている。
6393	中国漁船・尖閣領海内接触：船長処分、国際関係も考慮 政府が答弁書	政府は 19 日、沖縄県・尖閣諸島周辺で起きた中国漁船衝突事件に関し、容疑者を起訴するかどうかの刑事処分を検察官が判断するに当たっては「国際関係への影響などについても、犯罪後の状況として考慮できる」との答弁書を閣議決定した。自民党の高市早苗衆院議員の質問主意書に答えた。
6832	中国漁船・尖閣領海内接触：ビデオ流出 馬淵国交相、調査状況を仙谷長官に報告 【大阪】	仙谷由人官房長官は 7 日、首相官邸で馬淵澄夫国土交通相と会い、沖縄県・尖閣諸島沖での中国漁船衝突事件に関する映像流出問題の調査状況について報告を受けた。、 会談後、馬淵氏は記者団に「調査の報告をした」と話した。
6765	中国漁船・尖閣領海内接触：衝突映像、ネット流出か	沖縄県・尖閣諸島沖の中国漁船衝突事件のビデオ映像とみられる映像が、インターネット上に流出していることが分かった。映像は数種類あり、海上保安庁の巡視艇に、中国漁船が衝突している場面などが映っていた。海保は「コメントできない」としている。

表 3 提案手法のクエリ”中国漁船衝突事件”の初期検索結果

記事	タイトル	本文
6765	中国漁船・尖閣領海内接触：衝突映像、ネット流出か	沖縄県・尖閣諸島沖の中国漁船衝突事件のビデオ映像とみられる映像が、インターネット上に流出していることが分かった。映像は数種類あり、海上保安庁の巡視艇に、中国漁船が衝突している場面などが映っていた。海保は「コメントできない」としている。
7815	ことば：中国漁船衝突事件	中国漁船衝突事件、 沖縄・尖閣諸島付近で 9 月 7 日、中国漁船が海上保安庁の巡視艇に衝突。石垣海上保安部が公務執行妨害の疑いで中国人船長を逮捕した。中国側は反発し、船長の釈放を要求。レアアース（希土類）の輸出手続きの滞りや日本人拘束事件は、対日報復の可能性が指摘された。日本側は船長を釈放したが、中国で反日デモが相次いだ。11 月 4 日、衝突映像がインターネットの動画投稿サイトに流出。同 10 日には、神戸海上保安部の一色正春・元主任航海士（43）が流出を告白した。
6776	中国漁船・尖閣領海内接触：ビデオ流出 菅首相、原因究明指示	菅直人首相は 5 日午前の閣僚懇談会で、中国漁船衝突事件の映像の流出問題について、「しっかり調査して、原因究明しなければならない」と情報管理態勢の点検を指示した。
6832	中国漁船・尖閣領海内接触：ビデオ流出 馬淵国交相、調査状況を仙谷長官に報告 【大阪】	仙谷由人官房長官は 7 日、首相官邸で馬淵澄夫国土交通相と会い、沖縄県・尖閣諸島沖での中国漁船衝突事件に関する映像流出問題の調査状況について報告を受けた。、 会談後、馬淵氏は記者団に「調査の報告をした」と話した。
5773	中国漁船・尖閣領海内接触：「機会があれば中国側に説明」??NY で前原外相	【ニューヨーク山科武司】中国漁船と海上保安庁巡視船の衝突事件について、前原誠司外相は 21 日、「機会があれば中国側に日本の立場を説明したい」と語った。国連総会のため訪問したニューヨークで記者団に語った。…
6055	菅首相：尖閣は固有領土、説明の意向 ASEM に出発	菅直人首相は 3 日、ブリュッセルで開かれるアジア欧州会議（ASEM）首脳会議に出席するため、政府専用機で羽田空港を出発した。沖縄県・尖閣諸島周辺での中国漁船衝突事件をめぐる菅首相は 2 国間会談の場で、尖閣諸島は日本固有の領土とする立場を説明する意向だ。…
6923	中国漁船・尖閣領海内接触：ビデオ流出 中国も関係修復を模索 首脳会談へ努力要求	【北京・成沢健一】中国外務省の洪磊・副報道局長は 11 日の定例会見で、尖閣諸島沖衝突事件の映像を流出させたことと神戸の海上保安官が認めていることについて、「報道を注視している。中国側は、いわゆるビデオ問題が中日関係を妨げ続けることを望まない」と述べた。…

表 4 Rocchio アルゴリズムのクエリ”民主 大敗”の初期検索結果

記事	タイトル	本文
1663	日米密約：首相経験者らに国会出席呼びかけへ	民主、社民、国民新の与党 3 党の国対委員長は 10 日、国会内で会談し、日米両政府による外交密約の存在が明らかになったことを受け、歴代首相経験者や外相経験者に対し国会への出席を呼びかける方針で一致した。衆院外務委員会など衆参両院の関係委員会での参考人招致を想定している。【近藤大介】
3686	小沢・民主前幹事長：辞任タイミング、参院選前「ぎりぎりセーフ」	民主党の小沢一郎前幹事長は 12 日午前、連合和歌山の村上正次会長らと会談し、幹事長辞任について「もう少し早ければとも思ったが、鳩山由紀夫前首相と話し合っああいうタイミングになった」と説明。参院選を念頭に「ぎりぎりセーフかなと思う」と述べた。…
7551	11 年度予算：普天間予算、見送り要求??社民	社民党は 10 日午前、首相官邸で開かれた 11 年度予算編成を巡る民主、国民新両党との協議で、米軍普天間飛行場（沖縄県宜野湾市）の名護市辺野古崎地区への移転に関し、環境影響評価など関連経費の計上をすべて見送るよう求めた。…
1020	亀井金融・郵政担当相：「夫婦別姓」にも反対表明	国民新党代表の亀井静香金融・郵政担当相は 11 日、福井市内で講演し、民主、社民両党が導入を目指す永住外国人の地方選挙権と選択的夫婦別姓に反対する考えを示した。、亀井氏は夫婦別姓について、「一家の表札が（複数になって）アパートみたいになってしまう」と批判。「（この二つは）国民新党が反対するので、絶対に成立しない。法案提出も私がノーと言えられない」と話した。【大久保陽一】
1197	日本航空：過去の問題点を検証、民主が PT 発足 24 日にも初会合	民主党は 18 日、経営破綻（はたん）した日本航空の過去の問題点などを検証するプロジェクトチーム（PT）を発足させることを決めた。24 日にも初会合を開く。自民政権の実質的な保護下で粉飾決算がなかったかという問題や、空港建設に与党政治家がどう関与したかなどについても調査する。…
836	基幹労連：政治と金、民主の対応批判??内藤委員長	民主党を支持している「基幹労連」の内藤純朗委員長は 3 日、東京都内で開いた労連の中央委員会で「民主党の対応に一抹の不安と不信を抱くのは私だけではない」と、同党の政治と金にまつわる対応を批判した。…
1188	10 年度予算案：年度内成立へ	衆院は 18 日の本会議で、自民党が提出した鹿野道彦予算委員長（民主）の解任決議案を与党 3 党の反対多数で否決した。公明党とみんなの党は賛成し共産党は棄権した。予算委は 19 日に地方公聴会、24 日に中央公聴会が開かれることが決まり、10 年度予算案は 3 月 2 日までに衆院を通過する見通しになった。…

表 5 提案手法のクエリ”民主 大敗”の初期検索結果

記事	タイトル	本文
3686	小沢・民主前幹事長：辞任タイミング、参院選前「ぎりぎりセーフ」	民主党の小沢一郎前幹事長は 12 日午前、連合和歌山の村上正次会長らと会談し、幹事長辞任について「もう少し早ければとも思ったが、鳩山由紀夫前首相と話し合っああいうタイミングになった」と説明。参院選を念頭に「ぎりぎりセーフかなと思う」と述べた。…
4500	菅首相：小林・宮台両教授招き、参院選大敗の原因を聞く	励まされたり、反省したり、菅直人首相は 25 日、首相官邸に小林良彰慶応大教授（政治学）と宮台真司首都大学東京教授（社会学）を招き、参院選で民主党が大敗した原因について意見を聞いた。…
4731	菅首相：消費税率「10%」事実上撤回 「党政調で」	菅直人首相は 4 日夜、自らの「消費税率 10%」発言について記者団に「超党派の協議を呼びかけた中で自民党の案に触れたことがいろいろと誤解を生んだ。今は（民主）党の議論を待っている」と語った。…
4268	菅首相：消費税争点化で党に迷惑かけた??稲盛氏に	菅直人首相は 14 日午前、首相官邸で内閣特別顧問の稲盛和夫日航会長と会談し、参院選敗北について「唐突に消費税問題を挙げたからこういう結果になり、（民主）党全体に迷惑をかけた。大変反省している」と語った。…
194	山口・公明代表：「民主と連携、検討も」参院選後へ秋波、選挙協力にも含み	公明党の山口那津男代表＝似顔絵＝は 10 日、NHK の報道番組で、次期参院選後の民主と連携について「選挙結果に応じて対応を検討するのが国民への責任だ。国民のニーズを実現する視点で選択したい」と述べ、前向きな姿勢を示した。参院選での民主、自民両党との選挙協力に関しては「政党対政党の協力は今、考えていない」と否定する一方で、「人間関係中心」とも述べ、選挙区事情によっては民主党との協力にも含みを持たせた。【田所柳子】
4616	臨時国会：ヤワラちゃん、笑顔の初登院 【大阪】	参院選で 35 万票余りを獲得して当選した柔道家、谷亮子氏（34）＝民主、比例＝は 30 日午前 8 時 40 分ごろ、ページュのスーツ姿で初登院。多くの報道陣が集まり、警備にあたる衛視らともみ合いになる場面も。柔道との両立が可能か議論を呼んでいるが、意気込みを問われた谷氏は「公務をおろそかにすることなく務めたい」と話した。
3591	荒井国家戦略相：事務所費問題 自民・石破氏「辞任に値する」	自民党の石破茂政調会長は 9 日の記者会見で、荒井聡国家戦略担当相の事務所費問題に関し、「当然、辞任に値する。菅直人首相の任命責任も問われる」と辞任を求めた。

表 6 提案手法のクエリ”民主 大敗”のフィードバック 2 回目の検索結果

記事	タイトル	本文
4500	菅首相：小林・宮台両教授招き、参院選大敗の原因を聞く	励まされたり、反省したり、菅直人首相は 25 日、首相公邸に小林良彰慶応大教授（政治学）と宮台真司首都大学東京教授（社会学）を招き、参院選で民主党が大敗した原因について意見を聞いた。…
4268	菅首相：消費税争点化で党に迷惑かけた??稲盛氏に	菅直人首相は 14 日午前、首相官邸で内閣特別顧問の稲盛和夫日航会長と会談し、参院選敗北について「唐突に消費税問題を挙げたからこういう結果になり、（民主）党全体に迷惑をかけた。大変反省している」と語った。稲盛氏が記者団に明らかにした。首相は「政権運営は大変難しくなる。今後、慎重にする」とも語ったという。…
4320	菅首相：「任期 3 年、続投」意欲示す	菅直人首相は 16 日夜、東京都内のホテルで法政大学の五十嵐敬喜、江橋崇両教授と会食し、首相続投に強い意欲を示した。五十嵐氏によると、同氏らが「（衆院の残る任期の）3 年間、絶対に辞めないでください」と求め、首相は「任期まで全部やる」と述べ、今後の政権運営に強い意欲を示したという。一方、参院の過半数確保に関し、…
4731	菅首相：消費税率「10 %」事実上撤回 「党政調で」	菅直人首相は 4 日夜、自らの「消費税率 10 %」発言について記者団に「超党派の協議を呼びかけた中で自民党の案に触れたことがいろいろと誤解を生んだ。…
3686	小沢・民主前幹事長：辞任タイミング、参院選前「ぎりぎりセーフ」	民主党の小沢一郎前幹事長は 12 日午前、連合和歌山の村上正次会長らと会談し、幹事長辞任について「もう少し早ければとも思ったが、鳩山由紀夫前首相と話し合っあいうタイミングになった」と説明。参院選を念頭に「ぎりぎりセーフかなと思う」と述べた。、小沢氏は「政権がこういう形で終わり迷惑をかけた」と謝罪。今後の地方行脚については「幹事長の時のような動き方はできないが、ぼちぼちやる」と述べた。
4351	ファイル：民主・細野氏も予算委に前向き	民主党の細野豪志幹事長代理は 18 日のフジテレビ番組で、臨時国会について「野党側が求める予算委員会は少なくともやらないといけない」と述べ、参院で多数を握る野党側の主張を取り入れるべきだとの考えを示した。…
4235	枝野・民主幹事長：消費増税、年度内取りまとめ断念示唆	民主党の枝野幸男幹事長は 12 日の会見で、消費税増税に関する具体案の取りまとめについて「当初想定していた期限にこだわるのではなく、幅広い国民の理解と合意を得られるペースを進めることを基本に考えないといけない」と述べた。…