

言論マップ生成技術の現状と課題

水野 淳太^{†‡} Eric Nichols[‡] 渡邊 陽太郎[‡] 村上 浩司[§] 松吉 俊[†] 大木 環美[†] 乾 健太郎^{††} 松本 裕治[†]

[†] 奈良先端科学技術大学院大学

[‡] 東北大学

[§] 楽天技術研究所

1 はじめに

ウェブ上には大量のテキスト情報が存在し、そこでは様々なトピックに関して多角的な意見が述べられている。情報検索技術の発展により、あるトピックに関連する文書集合を容易に入手できるようになった。しかしながら、これらの文書に記述されている情報は、そのすべてが真実というわけではなく、不正確な記述、偏りのある意見などが混在している可能性が高い。そのため、あるトピックに対する情報の集合を俯瞰するには、ユーザは個々の情報の信憑性を判断する作業を繰り返すことを強いられる。しかし、限られた時間で各情報にそのような作業を行うことは容易ではない。これらの作業に関してユーザを支援する技術が必要である。

我々は現在、こうしたユーザによる Web 情報の信憑性分析を支援するために、言論マップ生成課題 [1] に取り組んでいる。これは、例えばユーザが「ディーゼル車は環境に良い」と思っていた場合、それをクエリとして入力すると、図 1 のようなクエリに関連する情報を提示する言論マップを出力するものである。ユーザに提示すべき情報は、〈同意〉、〈同意根拠〉、〈対立〉、〈対立根拠〉、〈弱対立〉、〈弱対立根拠〉の 6 種類であると考えている。こうした情報を中心として言明を整理することで、ユーザによる言明の信憑性判断の支援情報とする。

村上ら [2] では、いくつかのクエリに対して言論マップ生成を行い、それに対する考察を行った。その後、技術開発を続け、約 1000 文対に対する定量評価と、100 人規模のユーザ評価を行った。本稿では、局所構造アライメントや弱対立関係認識といった主要技術について現状を述べる。次に、システムの性能評価およびユーザ評価結果を報告し、エラー分析により今後どのような問題に着手していくべきかを論じる。

2 関連研究

ユーザによる情報の信頼性判断についての関連研究として、WISDOM[3] が挙げられる。WISDOM では、情報発信者の同定、意見分析といった技術で信頼性判断の支援を行っている。本研究では、賛成・反対といった二局的な意見だけでなく、条件付きで賛成・反対を表している意見の抽出や、それらの根拠となる情報の提示を行うことに注力している。

Dispute Finder[4] は、様々な情報に対する反論情報をデータベースに蓄積し、ユーザが閲覧中のウェブページ中で文を選択すると、その反論候補を提示するシステムである。ユーザがクエリを考えるのではなく、ウェブ上の情報の信憑性を即時に調査できる点が特長だが、対象としている関係が反論のみであることが本研究と大きく異なる。

文間の意味的関係の自動認識は、近年 NLP が実現すべき課題として精力的に研究され、いくつかのタスクが提案されてきた。その一つに、与えられた 1 組の文対が〈含意〉、〈矛盾〉もしくは〈不明〉のいずれであるかを判定

する課題として含意関係認識 (RTE)[5] がある。また、複数文書中の文間の関係解析には、CST (Cross-Document Structure Theory)[6] がある。RST[7] に基づく談話構造解析が単一文書内の構造を解析するのに対し、CST はこれを文書横断構造解析に拡張するものであり、18 種類の意味的関係が定義された。

CST では、Zhang らが比較する 2 文からのみ素性を抽出して素性空間に表現し、主要な関係を 1 つの分類器により分類を行った [8]。また、正解ラベルが付与されたデータが少ないためラベルなしデータも用いた Boosting を利用した手法を提案した [9] がどちらにおいても精度は高くなく、類義語や反義語等の語彙知識の適用、素性の洗練、各意味的関係認識のための個別処理の検討など、多くの課題を残した。このことから、複数の意味的関係クラスへの分類手法そのものの確立も重要な課題となる。

3 言論マップ生成システムの主要技術

言論マップの生成過程は、図 2 に示すように、大きくパッセージ検索と文間関係認識からなる。

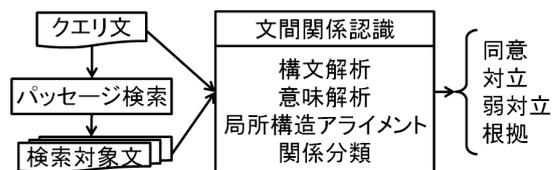


図 2: 言論マップ生成システムの流れ

入力には、「キシリトールは虫歯予防に効果的だ」のような、Yes/No で回答できるクエリ文を想定している。次に、クエリ文に対してパッセージ検索を行い、検索対象文を獲得する。これは、ウェブ検索エンジン²を利用することで、比較的容易に実現できる。従って言論マップ生成システムの主問題は文間関係認識である。この問題は、クエリ文と検索対象文との意味的関係を同定する課題である。より一般的には、1 組の文対が与えられたときに、その間の意味的関係を同定する課題であり、RTE はその部分問題に位置づけられる。文間関係認識のアプローチは、変形に基づく方法とアライメントに基づく方法に大きく分けられる [10]。変形に基づく方法は、受動態と能動態の変換や上位語と下位語の入れ替えなどの変形を行い、一方の文が他方の文に変形可能である場合に〈含意〉、それ以外の場合は〈その他〉に分類する。この方法は文対が〈含意〉関係にあるかどうかを判断することが中心であるため、多様な関係に分類するには不向きである。

そこで、本研究ではアライメントに基づく方法を採用する。この方法は、1) 2 文をそれぞれ意味解析、2) 文間で類似・関連する単語に対応付け (単語アライメント)、3) 1,2 の情報を利用して関係分類を行う。言論マップ生成システムでは〈同意〉、〈対立〉、〈その他〉の 3 種類への分類を行うが、Marneffe et al.[11] と同様に、まず〈その他〉であるかどうかを分類し、次に〈同意〉と〈対立〉に分類する。

¹ インターフェースでは簡単のため、〈同意〉→ 賛成意見、〈対立〉→ 反対意見、〈弱対立〉→ 一部反対としている

² 本研究では、TSUBAKI(<http://tsubaki.ixnlp.nii.ac.jp/>) を利用した

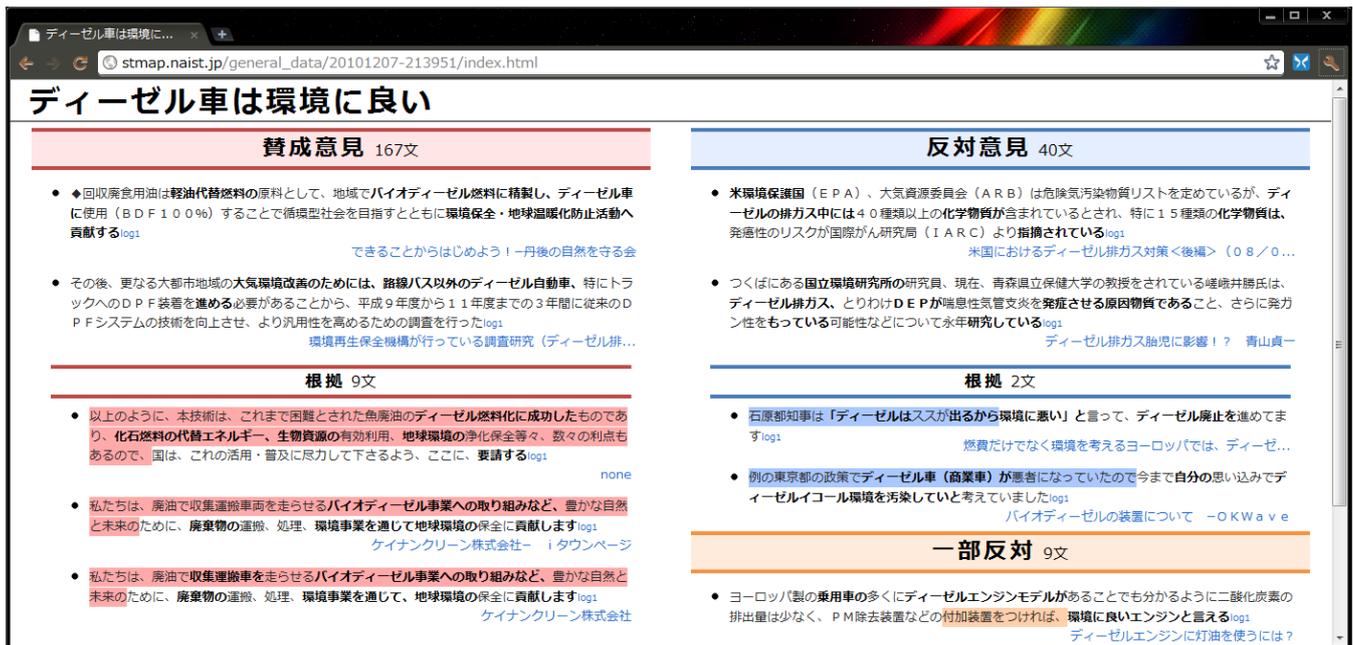


図 1: 言論マップ生成例¹

この文間関係認識については、これまで特に次の2つの技術的課題に注力してきた。一つは文間の単語アライメントにおいて、2単語間の意味的な関係について考慮する局所構造アライメントである。もう一つは〈弱対立〉、〈根拠〉という RTE にはない新しい関係への分類である。以下では、この2点に絞って技術開発の現状を報告する。

3.1 局所構造アライメント

単語アライメントとは、文間で類似・関連する単語間に対応付けを行うことで、関係分類の際に着目すべき部分を明らかにすることである [12][13] が、文の意味も考えると、類似単語であっても対応付けを行うべきではない場合がある。Chang et al. [14] は、クエリ側で依存関係にある2単語に対して、それが単語対応するテキスト側の2単語も依存関係にある場合のみ単語アライメントとして採用するという手法を RTE に適用した。しかしながら、 T_1 の「ブラックバス-破壊」のように、依存構造を持たない場合には有効ではない。 T_1 では、述語項構造解析によって「ブラックバスが破壊する」という格関係が成り立つことが分かり、 Q の依存構造と対応付けられる。次に、 T_2 では、同じように「魚類が破壊する」という格関係が成り立ち、かつ「ブラックバス」が「魚類」の例示関係にあるため、 Q の依存構造と対応付けられる。 T_1, T_2 のように、多くの場合は類似した単語に対応付けを行うことで、意味的な対応もとれている。しかし、 T_3 では「ブラックバス」と「破壊する」の間に格関係は成り立たず、 Q の依存構造と対応付けるべきではない。局所構造アライメントは、 $Q-T_1$ と $Q-T_2$ は対応させ、 $Q-T_3$ は対応させないようなアライメントである。

- (1) Q ブラックバス₀ は生態系₁ を破壊する₂
 T_1 ブラックバス₀ が増えたと生態系₁ が破壊される₂
 T_2 ブラックバス₀ のように癡猛な魚類が生態系₁ を破壊する₂
 T_3 ブラックバス₀ を駆除することが生態系₁ を破壊する₂

まず対応付ける単位について、英語を対象とした場合は単語単位での対応付けが一般的だが、日本語では「形態素」

が「文節」のいずれかである場合が多い。本研究では「文節」単位で対応付ける文節アライメントを採用するが、どちらがより良いアライメント単位であるかは今後の課題である。

局所構造アライメントは以下の手順で行う。

1. 文節アライメント
文間で類似する文節に対して対応付けを行う [15] .
2. ルールによる局所構造アライメント
クエリ側で依存構造を持つ2文節に対して、それが文節対応するテキスト側の2文節も依存構造を持つ場合に対応付けを行う。
3. 機械学習による局所構造アライメント
ルールで局所構造対応しなかった場合には、2組の文節を入力として、教師あり機械学習により局所構造対応するかどうかを二値分類する。ただし、あらゆる文節の組み合わせを考慮することは計算量の観点から現実的ではないため、クエリ側については依存構造を持つ2文節に限定する。

3.2 〈弱対立〉および〈根拠〉の認識

〈弱対立〉関係は、命題に対して部分的に〈同意〉や〈対立〉の関係にある文に対して付与される関係であり、命題が成立するための条件や、成立する程度や範囲についての制限といった情報を含んでいる [16] . 例えば、(2) では、下線部を考慮しない場合、〈同意〉の関係にある。しかし、下線部には条件が示されており、この条件が満たされなければ Q の命題が成り立たないことが示唆されている。

- (2) Q キシリトールは虫歯予防に効果的だ
 T キシリトールは 毎食後摂取して初めて 虫歯を予防できる 〈弱対立〉

〈根拠〉関係は、命題に対して〈同意〉や〈対立〉の関係であり、かつその根拠や理由を含む文に対して付与される関係である。例えば、(3) では、下線部を除けば〈対立〉の関係にあるが、下線部にはその根拠が示されている。

- (3) Q コラーゲンでシワが取れる
 T コラーゲンは肌から浸透しないので 出来てしまったしわ に対しての効果は期待できない 〈対立根拠〉

いずれの関係も、単純な〈同意〉、〈対立〉に比べて、ユーザにとって有用な情報が含まれており、それらを区別・強調して示すことは、命題の信憑性判断に対して重要な役割を果たすと考えられる。

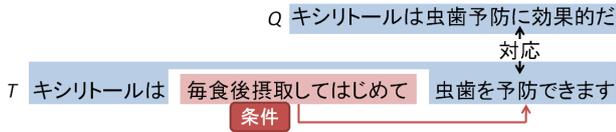


図 3: 弱対立関係の認識

〈弱対立〉および〈根拠〉認識の手法について、基本的なアプローチは共通であるため、〈弱対立〉認識を例に述べる。図 3 に示すように、局所構造アライメントの情報を元に、 T 中で Q と対応する部分を同定し、対応部分が〈同意〉または〈対立〉関係にある場合のみ、次の修飾関係の解析を行う。帰結部を修飾する文節について、条件表現や程度表現の語彙リストに含まれたら〈弱対立〉関係であると判断する [16]。〈根拠〉認識では、同様に帰結部を修飾する文節が理由や根拠を示すような語彙表現（「～ため」など）である場合に〈根拠〉関係であると判断する [17]。

4 評価実験

関係分類の性能評価（以下、性能評価）と、実際に言論マップ生成システムを利用したユーザによる主観評価（ユーザ評価）を行った。性能評価では、アライメントの違いに対する関係分類性能を比較することで、提案手法の有効性を示す。また、局所構造アライメントの正解データを用いることで、アライメント以外を要因とする問題について考察する。ユーザ評価では、評価者は言論マップ生成システムとウェブ検索エンジンの比較を行い、主観評価する。

4.1 実験設定

実験では、5 種類のクエリに対してウェブから獲得した 1326 文対を開発データとし、同様に 20 種類のクエリに対してウェブから獲得した 1050 文対を評価データとした。2 つのデータの間で、クエリの重なりは存在しない。評価データはクエリとの類似性に重みを置いたサンプリングを行ったため、〈同意〉関係にある文対の占める割合が多い。局所構造アライメントの正解は、開発データの全てと評価データのうち 1 割に対して付与し、関係ラベルの正解はすべてのデータに対して付与した。評価の対象とする関係は、〈同意〉、〈対立〉、〈その他〉の 3 種類で、上記の通り〈同意〉の割合が高いため、他の 2 関係への分類性能が重要である。

4.2 性能評価

性能評価では、以下の 3 種類の手法の比較を行う。

ベースライン 1 文構造を考慮せず、文節アライメント結果のみを利用

ベースライン 2 ルールによる局所構造アライメント結果を利用

提案手法 機械学習による局所構造アライメント結果も利用
局所構造アライメントは、開発データで学習し、評価データに対してシステム出力を作成した。関係分類は、開発データを常に学習に用い、評価データに対して、学習データとテストデータの間で同一クエリが重複しないように 10 分割交差検定を行った。学習の際に、評価データ内の 1 割のアライメント正解データも用いた。

実験結果を表 1 に示す。# は事例数、Pre は精度、Rec は再現率を示している。全体として、ベースライン 1 およ

び 2 と比較して提案手法が上回ったが、その差は限定的であった。誤り分析を行った結果、その多くはアライメントの誤りが原因であることが分かった。

そこで、関係分類の学習およびテストにアライメントの正解データを用いることで、上限性能を調べた。実験結果を表 1 に示す。上限性能の結果については、アライメントの正解データは全体のうちの一部にしか付与されていないため、他の実験とはデータ総数が異なる。全ての関係において、ベースラインや提案手法に比べて精度、再現率ともに大幅に向上している。この結果から、文間関係認識におけるアライメントの重要性が確認できた。

しかし、アライメントの正解データを用いても、未だ正しく関係を分類できていない事例が残っている。次節にて、上限性能での実験結果における誤り分析について述べる。

4.3 誤り分析

アライメントの正解データを用いた実験の結果、正しく関係を分類できなかった事例は 102 事例のうち 24 事例であった。これらのうち、主要な誤りについて以下で例を用いながら述べる。

否定、反義: 否定、反義に起因する誤りが 4 事例あった。

(4) Q コラーゲンは肌に良い

T コラーゲンが減少すると肌のハリが失われます
(正解:〈同意〉システム:〈対立〉)

Q と T において、「肌に良い」と「肌のハリが失われる」の意味の関係は対立であるが、「減少」のように負の極性、否定的な表現が入ると意味の関係は逆転する。これらを同意関係として扱いたい。そこで、現状では「減少」や「低下」など、意味的关系を逆転させるような単語のリストを作成し、分類時の手がかりとして用いている。しかし、規模がまだ小さく網羅性に欠けるため、認識に失敗しているさらなる拡充が必要である。

含意・叙実述語: 含意、叙実述語を含む誤りが 3 事例あった。

(5) Q 酔を飲むと身体が柔らかくなる

T 「あんなに大量の酔を飲むから、サーカス団員は身体が柔らかい」と噂したことから生じた誤解である
(正解:〈対立〉システム:〈同意〉)

含意、叙実述語は、文間関係認識技術の実現において重要な役割を果たす。上記の例において、「誤解」は反叙実動詞の名詞化形であり、「A は誤解である」という言明から「A でない」が含意される。したがって、「誤解」という単語が上記のような含意関係のトリガとなることを捉える必要性があるが、現在の関係分類では対応できていない。

構文: 逆接構文や比較構文を含む誤り事例が 3 事例あった。逆接構文は、例えば、「ミネラルウォーターは水道水より安全だと言われているが、それは誤りである」のように、前方において完全に対応関係があるが、後方での否定により、文全体として対立関係になるような事例である。比較構文は、構造的類似度に基づく関係認識手法では認識が難しい事例の一つである。

(6) Q ミネラルウォーターは水道水より安全だ

T 水質基準だけ見れば、ミネラルウォーターよりも厳密に管理された水道水の方が安全である
(正解:〈対立〉システム:〈同意〉)

上記の例の場合、 Q と T の対応付けだけでは、正しく関係を分類することは困難である。 Q と T の双方に比較構文が存在するため、それぞれにおいて比較対象を認識し、

表 1: 性能評価実験の結果

	#	同意 F1 (Pre, Rec)	対立 F1 (Pre, Rec)	その他 F1 (Pre, Rec)	正解率 (%)
ベースライン 1	921	72.68 (69.42, 76.25)	37.53 (34.27, 41.46)	19.87 (28.85, 15.15)	55.37
ベースライン 2	921	72.91 (70.18, 75.87)	37.80 (34.40, 41.95)	20.71 (28.83, 16.16)	55.48
提案手法	921	72.74 (72.33, 73.17)	32.90 (35.39, 30.37)	33.09 (31.51, 34.85)	55.48
上限性能	102	80.43 (75.51, 86.05)	74.58 (84.62, 66.67)	71.70 (70.37, 73.08)	76.47

それらの間の関係が Q, T でどのように述べられているかを認識する必要がある。これは、構造的類似度に基づくアプローチとは異なる枠組が必要である。

4.4 ユーザ評価

ユーザによる 言論マップ生成システムとウェブ検索エンジン³との比較実験・主観評価を行った。

クエリは 3 種類のカテゴリからなる 54 種類を用意した。カテゴリとクエリの一例を以下に示す。

社会・生活問題 「裁判員になるのを拒否できる」「血液型で性格が分かる」

健康問題 「ミネラルウォーターは水道水より安全だ」「アガリクスは健康に良い」

環境問題 「南極の氷は減っている」「地球温暖化によって海面が上昇する」

作業者は 112 名で、54 クエリの中から 4 クエリを自身の興味に従って選択し、評価を行った。実験設定の詳細については、文献 [18] を参照されたい。

ユーザが試行する作業は、54 種類のクエリの中から任意に選択したクエリに対して、1) ウェブ上に存在する同意、対立意見、2) その根拠となる情報を見つける、というものである。2 つの作業を言論マップ生成システムと検索エンジンの両方で行い、主観評価を行った結果、ウェブ検索エンジンよりも優れた評価が得られた。

アンケートには、ウェブ検索エンジンと比較して「物事を様々な視点から見るのに役立つ」といった意見が多く、ユーザに信憑性判断の支援を行うという本研究の意図が正しく受け入れられていることが分かった。否定的な意見の多くは、インターフェースのデザインや分類精度の低さを指摘するものであり、分類性能の改善は当然だが、インターフェースにも注力していく必要があることが分かった。特に、検索対象文中の情報についてさらに調査を行いたい場合、ウェブ検索エンジンではその情報をクエリ単語として新たな検索が可能だが、本システムではクエリ文を作成する必要がある。今後、検索対象文の一部を選択するだけで自動的にクエリを生成するなど、より深い調査を容易に行えるようにしていくことが考えられる。

5 おわりに

本稿では言論マップ生成システムについて、その主要技術である局所構造アライメントや弱対立関係認識について、その現状を述べた。評価実験では、約 1000 文対を対象とした関係分類性能評価を行うことで、アライメントの重要性を確認した。また、アライメントの正解データを用いた実験により、叙実述語や比較後文への対応が必要であることが明らかになった。ユーザ評価では、本研究の意図がユーザに正しく受け入れられていることが確認でき、ウェブ検索エンジンと比較して高い評価を得られた。

今後は、エラー分析に基づく関係分類の性能向上や、インターフェースの充実による言論マップ生成システムの利便性向上などが課題である。

謝辞 本研究は (独) 情報通信研究機構の委託研究「電気通信サービスにおける情報信憑性検証技術に関する研究開発」の一環として実施した。

参考文献

- [1] Koji Murakami, Eric Nichols, Suguru Matsuyoshi, Asuka Sumida, Shouko Masuda, Kentaro Inui, and Yuji Matsumoto. Statement Map: Assisting Information Credibility Analysis by Visualizing Arguments. In *Proc. of WICOW 2009*, pp. 43–50, 2009.
- [2] 村上浩司, 水野淳太, 後藤隼人, 大木環美, 松吉俊, 乾健太郎, 松本裕治. 文間意味の関係認識による言論マップ生成. 言語処理学会第 16 回年次大会発表論文集 PA2-22, 2010.
- [3] Susumu Akamine, Daisuke Kawahara, Yoshikiyo Kato, Tetsuji Nakagawa, Kentaro Inui, Sadao Kurohashi, and Yutaka Kidawara. WISDOM: A Web Information Credibility Analysis System. In *Proc. of the ACL-IJCNLP 2009 Software Demonstrations*, pp. 1–4, 2009.
- [4] Rob Ennals, Beth Trushkowsky, and John Mark Agosta. Highlighting Disputed Claims on the Web. In *Proc. WWW 2010*.
- [5] Ido Dagan, Oren Glickman, and Bernardo Magnini. The pascal recognising textual entailment challenge. In *Proceedings of the PASCAL Challenges Workshop on RTE*, 2005.
- [6] Dragomir R. Radev. Common Theory of Information Fusion from Multiple Text Sources Step One: Cross-Document Structure. In *Proc. of the 1st SIGdial workshop on Discourse and dialogue*, pp. 74–83, 2000.
- [7] Mann William and Sandra Thompson. Rhetorical structure theory: towards a functional theory of text organization. *Text*, Vol. 8, No. 3, pp. 243–281, 1988.
- [8] Zhu Zhang, Jahna Otterbacher, and Dragomir Radev. Learning cross-document structural relationships using boosting. In *CIKM '03*, pp. 124–130, 2003.
- [9] Zhu Zhang and Dragomir Radev. Combining labeled and unlabeled data for learning cross-document structural relationships. In *IJCNLP '05*, pp. 32–41, 2005.
- [10] Shachar Mirkin, Ido Dagan, and Sebastian Padó. Assessing the role of discourse references in entailment inference. In *Proc. of ACL 2010*, 2010.
- [11] Marie-Catherine de Marneffe, Anna N. Rafferty, and Christopher D. Manning. Finding contradictions in text. In *Proc. of ACL 2008*, pp. 1039–1047, 2008.
- [12] Bill MacCartney, Michel Galley, and Christopher D. Manning. A Phrase-Based Alignment Model for Natural Language Inference. In *Proc. of EMNLP 2008*, pp. 802–811, 2008.
- [13] Mark Sammons, V. G. Vinod Vydiswaran, Tim Vieira, Nikhil Johri, Ming-Wei Chang, Dan Goldwasser, Vivek Srikumar, Gourab Kundu, Yuancheng Tu, Kevin Small, Joshua Rule, Quang Do, and Dan Roth. Relation Alignment for Textual Entailment Recognition. In *Proc. of Recognizing Textual Entailment 2009*, 2009.
- [14] Ming-Wei Chang, Dan Goldwasser, Dan Roth, and Vivek Srikumar. Discriminative learning over constrained latent representations. In *NAACL-HLT*, Los Angeles, California, 2010.
- [15] 水野淳太, 後藤隼人, 渡邊陽太郎, 村上浩司, 乾健太郎, 松本裕治. 文間関係認識のための局所構造アライメント. 情報処理学会研究報告 2010-NL-196, pp. 55–60, 2010.
- [16] Megumi Ohki, Eric Nichols, Suguru Matsuyoshi, Koji Murakami, Junta Mizuno, Masuda Shouko, Kentaro Inui, and Yuji Matsumoto. Recognizing confinement in web texts. In *Proc. of IWCS 2011*, 2011.
- [17] Eric Nichols, Junta Mizuno, Yotaro Watanabe, and Kentaro Inui. Toward evidence search. 言語処理学会 第 17 回年次大会, 2011.
- [18] 岡嶋穰, 河合剛巨, 中澤聡, 村上浩司, 松吉俊, 水野淳太, エリックニコルズ, 渡邊陽太郎, 乾健太郎, 洪木英潔, 中野正寛, 宮崎林太郎, 石下円香, 森辰則. Web 文書の時間・論理関係分析に基づく情報信頼性判断支援システムの開発と実証実験. 言語処理学会 第 17 回年次大会, 2011.

³本研究では、google(<http://www.google.co.jp>) を用いた