

大学入試化学の計算問題の自動解答システム

吉田達平*1 松崎 拓也*1 佐藤 理史*1
 Tappei Yoshida Takuya Matsuzaki Satoshi Sato

*1 名古屋大学大学院 工学研究科 電子情報システム専攻

We developed an automatic solver for high-school chemistry. First, the solver translates a problem to an intermediate representation that expresses the transition of chemical states in the form of a time series. Next, the solver interpolates the intermediate representation with additional information that is not explicit in the problem sentences. The additional information is supplied by several means including the retrieval of chemical knowledge from a database and language analysis such as coreference resolution. Finally, the solver calculates answer based on the fully-interpolated representation.

1. はじめに

本研究ではセンター試験「化学」の計算問題に取り組む。本研究は、国立情報学研究所を中心とする「ロボットは東大に入れるか」プロジェクトの一環である [新井 12]。化学の入試問題は、言語理解、化学知識の検索と利用、数学的推論の統合を必要とする複雑な質問応答課題である。さらに、化学の計算問題の主たる対象は、時間によって化学的な状態が変化する動的な世界であり、問題文からこの世界の変化を読み取るという言語処理における重要な課題を含む。本研究は、このような新しい質問応答の課題として化学問題の自動解答に取り組むものである。

化学の問題は、質問内容を特定しテキストからそれに関する記述を正しく検索する、という従来の質問応答の問題の枠組みだけでは解けない。化学の問題では、問題ごとに記述された内容を時間軸の形をした構造的なデータとして抽出し、このデータの各イベントについて背後にある化学的知識を検索し、その上で計算する必要がある。つまり、化学の計算問題には、化学的なイベントを経て状態が遷移していく世界を自然言語から抽出し、その上で演繹を行なうという大きな課題がある。「テキストから事実、知識を検索する」というのは、この言語処理から演繹処理の一連の処理の中のただ一要素にすぎない。これと対照的に、センター試験の「日本史」「世界史」では従来研究されてきた factoid 型の問題と true-or-false 型の問題が全体の 80~90% を占め [Miyao and Kawazoe 13]、これらのほとんどはテキストから適切な箇所を検索することで解けると考えられる。これは、歴史の問題のほとんどは、年号や年代によって時間軸にアンカー可能な、単一のイベントに関する質問で、状態の変化についての認識・推論を本質的に要さないためである。

化学の中でも特に計算問題に取り組む理由は 2 つある。1 つ目は計算問題は近年のセンター試験で問題全体の 1/5~1/3 と、比較的小おおくの部分を占めることである。2 つ目は化学の計算問題を観察した結果、ほとんどの問題が統一フォーマットで表現できることが分かったためである (次章で詳しく説明)。この 2 点から、化学の問題の中で効率よく点数を取ることができ、比較的易しいであろう計算問題に着目した。

化学の試験問題を計算機で取り扱った先行研究として、AP Chemistry という大学課程の化学の問題を対象として取り組

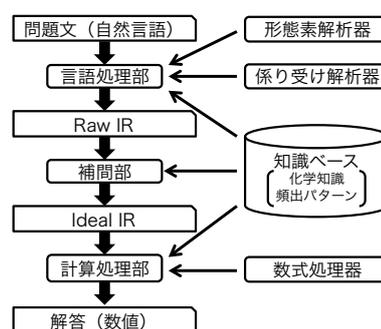


図 1: システムの全体像

んだものがある [Barker 04]。これは化学に関するデータベースの知識表現の方法に着眼し研究したものであり、言語処理には注力していない。本研究は、計算問題に限ったものであるが、自然言語を意味表現へ変換する言語処理を含め、化学の問題を解く一般的な枠組みを設計し計算機上で実現する方法を研究した。

2. システム

システムは言語処理部と計算処理部と補間部の 3 つで構成される。言語処理部は自然言語の問題文を入力として、中間表現を出力する。計算処理部は中間表現を入力として、方程式の立式・求解を行い解答を出力する。言語処理の結果として直接得られる中間表現と、計算処理の入力として求められる理想的な中間表現は別のものであり、これらをそれぞれ RawIR と IdealIR と呼ぶ。補間部はこの 2 つの中間表現のギャップを埋める処理を行なう。システムの全体像を図 1 に示す。

2.1 言語処理とその出力の中間表現

問題文中の化学用語にラベルを付与した後、問題文で表現された物質やエネルギー、化学的なイベント等に関する情報を抽出して構造化し、これを中間表現とする。問題文から中間表現への変換処理では、計算処理の手順は差し置き、言語表現とその構造のみに従って、単語や句の持つ情報を段階的にまとめて構造化することで情報を抽出する。

2.1.1 用語へのラベル付け

まずは、抽出すべき情報を持っている問題文中の単語にラベル付けをする。ラベル付けには 2 つの目的がある。1 つ目は、中間表現に含める情報として抽出すべき単語とそうでない単語を区別することである。2 つ目は、形態素解析の誤りを防ぐためである。ラベル付け対象の単語の多くを占める、物質名など

連絡先: 吉田達平, 名古屋大学大学院 工学研究科 電子情報システム専攻, 〒464-8603 名古屋市千種区不老町 IB 電子情報館南館 1 階 159 号室, 052-789-4435, 052-789-3146, tappei_y@nuce.nagoya-u.ac.jp

ラベル	単語
SUBNAME	塩酸, 水酸化ナトリウム水溶液, アルミニウム, ...
BODY	気体, 固体, 沈殿, ...
VERB	加える, 溶かす, つくる, ...
EVE	燃焼(する), 加熱(する), 溶解(する), ...
HEAT	熱量, 生成熱, 燃焼熱, ...
QUA	物質質量, 質量, 濃度, ...
PLENTY	十分
STANDARDSTATE	標準状態

表 1: 用語分類ラベルの例

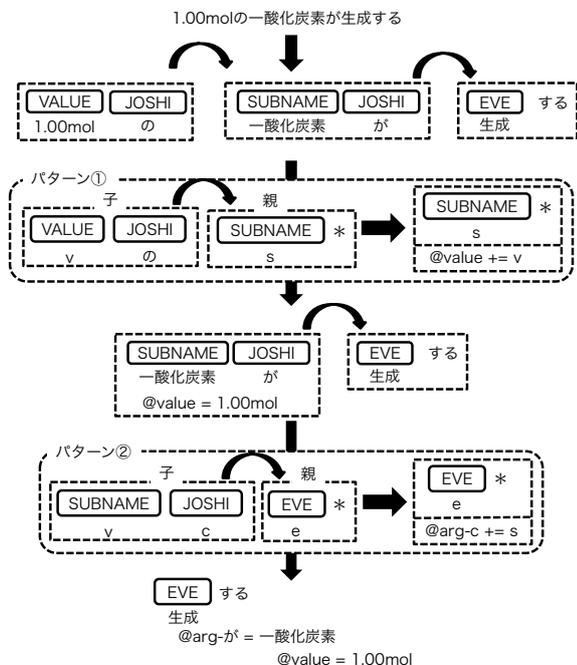


図 2: あるフレーズが意味構造に変換される例

の化学の用語は、辞書に登録されていないことが多く、また複数の形態素から構成されることが多いため、高い頻度で形態素解析誤りが起こる。表 1 に例を載せた。

2.1.2 パターンの段階的適用による情報の構造化

ラベル付の次に係り受け解析とパターンマッチによって意味構造を導出する。係り受け解析には cabocha を用いた [工藤 02]。パターンはセンター試験の化学の問題に頻出するフレーズを人手で抽出し一般化したものである。フレーズは 1 文節から係り受け関係にある 3 文節までの長さがある。すべてのパターンは 2 つ以上の単語を持ち、「親」と「子」の単語に分かれる。この区別は意味の構造化で利用する (後述)。

パターンがマッチした時にフレーズが持つ情報を部分的に構造化し、これを繰り返し完全な意味構造が得られる。部分的な構造化とはパターンの「子」の単語が持つ情報を、「親」の単語の属性として格納する操作である。図 2 に自然言語文から、意味構造に変換する例を示す。この例では、「1.00mol の → 一酸化炭素」というフレーズが、「VALUE (数値) + JOSHI (助詞) → SUBNAME (物質名)」というパターン ① に当てはまり、親である「一酸化炭素」の物理量に関わる属性の一つとして「1.00mol」が「一酸化炭素」の一つ下のノードに格納される。同様に「一酸化炭素が → 生成」というフレーズは、パターン ② に当てはまり「生成」の反応物質に関わる属性の一つとして「一酸化炭素 (が)」が格納されている。

上記の一連の処理の結果、意味構造を表す木の頂点になる単語を root と呼び、root のラベルの種類によってフレーズが持つ情報が分類できる。これは現状のシステムでは 4 種類あ

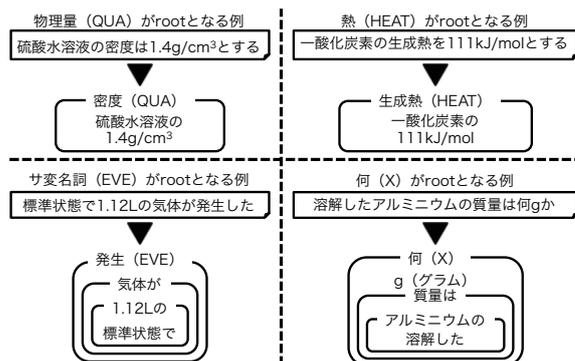


図 3: 4 種類のラベルの単語が root になるフレーズの例

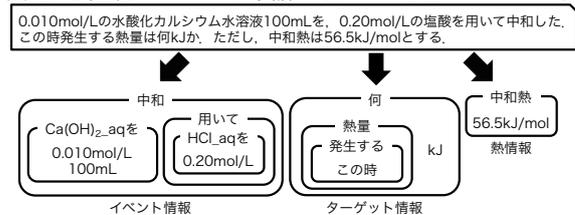


図 4: 1 問から得られた RawIR の例

り、これらのラベルそれぞれが実際に root となり構造化される文例を図 3 に載せる。この 4 種類のラベルが root になった場合、フレーズはそれぞれ次の意味を持つ。(1) **QUA (物理量)**: QUA の例は「質量」や「密度」や「物質質量」である。これが root になるフレーズは、質量や密度など物質の物理量の情報を与える。このタイプの情報を「物質情報」と呼ぶ。(2) **HEAT (熱)**: HEAT の例は「熱量」や「生成熱」や「溶解熱」である。これが root になるフレーズは、問題文中のイベントで発生した熱や、一定量の物質がある反応をした場合に一般的に発生する熱の情報を与える。このタイプの情報を「熱情報」と呼ぶ。(3) **EVE (サ変名詞) / VERB (動詞)**: EVE の例は「加熱(する)」や「分解(する)」であり、VERB の例は「加える」や「溶かす」である。これらが root になるフレーズは、イベントの発生や、そのイベントに関わった物質やその時の発生した熱などの情報を与える。このタイプの情報を「イベント情報」と呼ぶ。(4) **X (何)**: X は求めるべき数値を象徴する単語であり、過去問に登場したのは「何」のみである。主に「何 + [単位]」という形でフレーズ中に現れる。これが root になるフレーズは、その問題で求めるべき数値が何かを指定する。このタイプの情報を「ターゲット情報」と呼ぶ。

問題文全体を言語処理すると、一つ、または複数の意味構造木が得られる。この意味構造木の集合を RawIR とする。図 4 に 1 つの問題から得られた RawIR の例を示す。この問題では、イベント情報、ターゲット情報、熱情報の 3 つの意味構造木が得られる。

2.2 計算処理とその入力の中間表現

計算処理部の入力として理想的な中間表現 IdealIR は、求めるべき数値を付与した時系列の形で設計し、実際の過去問から人手で生成した例を図 5 に示す。時系列は状態と状態がイベントで結ばれた構造を持ち、時系列上の状態とイベントには問題文中で与えられた物理量などの情報と、求めるべき物理量を記述しておく。本研究ではこの時系列を timeline と呼び、求めるべき数値を target と呼ぶ。

計算処理は問題文で与えられていない timeline 上の数値を全て変数とし、それらの変数や定数の関係を表す連立方程式を立式し、これを解くことで行なう。例えば、図 5 の State1

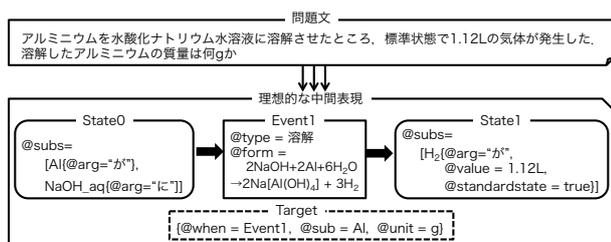


図 5: 理想的な中間表現の例

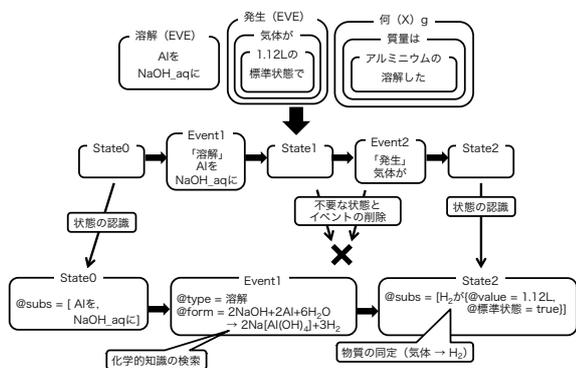


図 6: RawIR から IdealIR への補間の例

ではアルミニウムの体積や質量が与えられていないため、これらに変数を代入する。次に、計算公式に問題文で与えられた数値や、システムが与えた変数を代入し、方程式を立式する。公式の例には「密度 = 質量 / 体積」「(標準状態において) 体積 (L) = 物質質量 × 22.4」がある。timeline の情報と知識ベース内の公式を用いて生成しうる全ての方程式を数式処理器に入力し、target のラベルがついた変数の値を求める。

2.3 言語処理の出力と計算処理の入力の間の中間表現の補間

2.1 節で説明した言語処理部から出力される中間表現 RawIR と、2.2 節で説明した計算処理部に入力する中間表現 IdealIR の、両者の間のギャップを埋めなくてはならない。本稿ではこの処理を補間と呼ぶ。補間は問題文には直接かかれていないが、化学の知識や、日本語の常識から問題文の意味として推定すべき情報を抽出しているのであり、人間の受験生とまさに同じことをしていると言える。図 6 に RawIR から IdealIR への補間の例の一部分を示す。以下では図 6 を利用して補間の処理の流れを説明する。

timeline の作成: RawIR の情報を時系列に従って再構成する。まず、RawIR で得たイベント情報の集合 e_1, e_2, \dots, e_n の間と両端に状態 s_0, s_1, \dots, s_n を配置し、状態とイベントの配列 $s_0 \rightarrow e_1 \rightarrow s_1 \rightarrow \dots \rightarrow e_n \rightarrow s_n$ を仮の timeline とする。図 6 では、2 つのイベント「溶解」と「発生」と、その前後に 3 つの状態 State1, State2, State3 を作成して timeline とした。しかし「State2」と「State3」は化学的には同一の状態なので、図 6 にあるように、不要な要素として削除するべきである。これについては後述する。

次に、RawIR のターゲット情報では求めるべき物理量が抽出されているが、その timeline 上での場所を推定する必要がある。推定の方法には問題によって異なるいくつかの種類が考えられる。図 6 の問題では「溶解したアルミニウム」と書かれているので、「溶解」のイベントで反応したアルミニウムの質量と分かる。

状態の認識: イベントの前後に状態を作成して timeline を作成したが、状態の詳細は「物質情報」や「イベント情報」から

推定する必要がある。例えば、図 6 の問題では、「アルミニウムを溶解させた」という記述より、日本語の常識を以って「溶解」のイベントの前の状態に「アルミニウム」があると推定する。一方、「気体が発生した」という記述から「発生」のイベントの後の状態に「気体」が含まれる事が分かる。イベントの前か後かの判断は (1) 物質につく助詞 (2) イベントを象徴する動詞 / サ変名詞、の 2 つの要素の組み合わせで決定する。状態の認識は後述の、timeline 上のイベントに対応する化学反応式を検索する際に重要となる。

化学的知識の検索: 大学入試化学は知識の暗記とその知識を正しく利用することが重要で、計算機で解く場合にも同じである。例えば図 6 では「アルミニウム (Al) を水酸化ナトリウム水溶液 (NaOH, H₂O) に溶解」する場合、Al, NaOH, H₂O が項となる反応式を検索する。timeline 上のイベントに対応する反応式の特定は、計算処理や、次に述べる timeline から不要な要素を削除する処理で必要となる。

不要な状態とイベントの削除: 問題文中の動詞とサ変名詞を、全て timeline のイベントとするのは適切ではない。例えば図 6 で「溶解させたところ気体が発生する」場合、「溶解」は化学的な変化を伴うので timeline 上で「溶解」イベントの前後には別の状態を定義する必要がある。一方、「気体の発生」は「溶解」に付随して起こる物理的な現象で、「発生」前後で状態は化学的には変化しない。よって正しい timeline を得るには「溶解」のイベントと、その前の状態「State2」は timeline から削除する必要がある。図 6 では、「溶解」イベントの生成物の一つが「水素」と分かっているので、水素が「State3」に存在すれば、「State1」と「State3」が「溶解」イベントを経た遷移関係にあると推定でき、「State2」と「発生」イベントを「timeline」から削除することが可能になる。この問題に関しては、「State3」の物質が「気体」なので、後述の「物質の同定」において「気体」と「水素」が同一であることと複合的に推定する。この推定は問題によって複数のパターンが考えられる。

物質の同定: 物質は常に具体的な物質名で与えられるとは限られず、抽象的に表現された物質に対応する化学式を推定する必要がある。例えば、図 6 で「水素」が「気体」と書かれているように、表 1 の BODY のラベルに含まれる単語で与えられる場合が代表的である。この場合は、ある物質が標準状態や、常温常圧の水溶液中でどんな状態であるかをデータとして持ち、これを参照するという方法で対応する。この方法で「気体」と「水素」が同一のものであり、「水素」は「State3」に存在すると推定できる。さらに、前述の「不要な状態とイベントの削除」と本項「物質の同定」との複合的な推定により、「State2」と「発生」イベントを「timeline」から削除することが可能になる。

3. システムの現状

本稿では、開発データとしてセンター試験 6 回分の問題のうち、計算問題 26 問 74 文を用いた。この 26 問は 108 文からなるが、このうち中間表現として抽出すべき情報を含む文だけを対象とした。抽出すべき情報がない文とは、例えば単なる受験生に対する指示や、ヒントにはなっているが具体的な情報がなく中間表現として残らない文である。開発データ全体には計算問題は 28 問あるが、うち 2 問は図を含む問題であったため対象外とした。

システムが現在、問題文から解答の数値を導出できる問題は 26 問中 4 問である (図 6 の問題も含む)。解ける問題が少

ないのは、2章で説明し実装したシステムはシステムの計算問題に共通する骨組みであり、これのみで表現できるのはごく基本的な問題に限られるためである。今後、熱化学や電気分解の特別な処理が必要な問題や、次章で説明する問題を解決する個別的な要素を、本稿の中間表現とソルバーに追加する必要がある。処理の各ステップについては、(1) 言語処理部が問題文から RawIR に変換できたものは全 74 文中 22 文、(2) 補間部が RawIR から IdealIR に変換できたものは全 26 問中 4 問である、(3) 計算処理部が人手で書いた理想的な中間表現 IdealIR から解答が導出できたものは全 26 問中 10 問である。

4. 今後の課題

現在、言語処理できない文、補間・計算できない問題のうち、代表的な例をまとめる。

照応問題：大学入試の化学には言語知識のみで解決できるものから、化学的な知識を要求するものまで様々なパターンの照応問題が含まれる。例えば (1) 「このとき、陰極において銀が 0.015mol 析出した。」という記述において、「このとき」とは timeline 上でどこを指しているのか。(2) 「0.15mol/L 水酸化ナトリウム水溶液 140mL に、同じ温度の 0.075mol/L 硫酸水溶液 140mL を加える」という記述において、「同じ温度の硫酸水溶液」は何と同じ温度なのか。このように照応表現に関しては様々な問題が残されている。

多様な timeline の構造への対応：2章までは timeline は状態とイベントが交互に繋がった一本の鎖状の時系列構造を想定していたが、センター過去問にはそうでない問題が出題されることがあり、このような問題に対応する必要がある。具体的には以下の 3 パターンを確認している。以下に例を載せ図 7 にイメージを載せた

- 1. timeline が枝分かれするタイプ**
濃度未知のシュウ酸水溶液 A 25mL に十分な量の硫酸水溶液を加えて、0.050mol/L の過マンガン酸カリウム水溶液で滴定すると、過マンガン酸カリウムによる薄い赤色が消えなくなるまでに 20mL を要した。このシュウ酸水溶液 A 25mL を過不足なく中和するには、0.25mol/L 水酸化ナトリウム水溶液が何 mL 必要か。
- 2. timeline が平行するタイプ**
A 固体の水酸化ナトリウム 0.200g を 0.1mol/L の塩酸 100ml に溶かしたところ、505J の発熱があった。B 固体の水酸化ナトリウム 0.200g を水 100ml に溶かしたところ、225J の発熱があった。実験 (A・B) の結果から求められる、次の熱化学方程式の Q の値として最も適当な数値を求めよ。 $\text{HCl(aq)} + \text{NaOH(aq)} = \text{NaCl(aq)} + \text{H}_2\text{O} + \text{Q[kJ]}$
- 3. 複数の timeline が干渉しあうタイプ**
メタノール 64g を完全燃焼させて、20 °C の水 1.0kg を加熱する。発生する熱の 10% が、この水の温度上昇に使われるとすると、水の温度は何度になるか。

具体性が高過ぎる文の処理：センター化学の計算問題はほとんどが実験操作のような状況を題材にした問題であるが、表現が具体的になるほど言語処理が難しくなる。化学の計算問題を解くには、(1) 発生したイベント (2) イベントの要素となる物質 (3) 物質の物理量、この 3 つを正しく抽出した timeline を作成すればよい。具体的な実験手順などの記述によって、それ以外の情報がたくさん載るほど言語処理は難しくなる傾向がある。例えば、「ある量の気体のアンモニアを入れた容器に 0.30mol/L の硫酸 40mL を加え、よく振ってアンモニアを全て吸収させた」という文から抽出する必要がある情報の一つは、「加える (アンモニアに、硫酸を)」のようなイベント情報であるが、これを得るには 2.1 節に加えて別の処理が必要にな

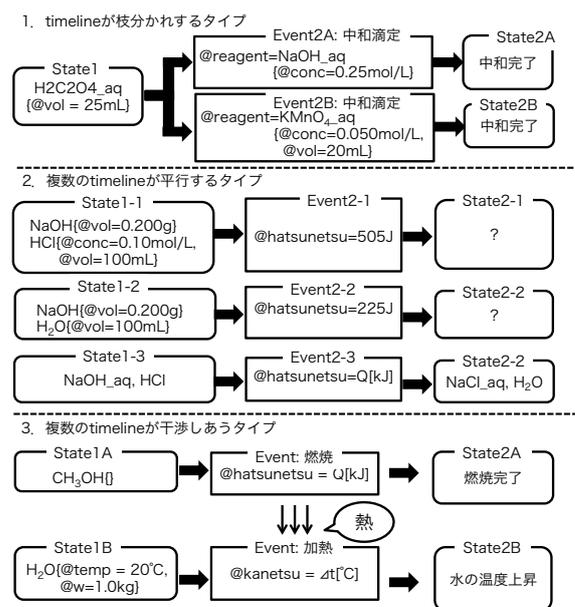


図 7: RawIR から IR への補間の例

る。なぜなら、単に 2.1 節で示した言語処理をすると「加える (容器に、硫酸を)」というイベントが抽出されるためである。「[物質] を→入れた→容器に→ [動詞]」というパターンを、「[物質] に→ [動詞]」と同一視するアドホックなルールを用いれば、このフレーズは処理できる。しかし、頑健なシステムを実現するには、同じ問題にグルーピングできるフレーズを多数収集するとともに、一般的な処理を講じる必要がある。

5. おわりに

センター試験「化学」の計算問題に対するソルバを設計、実装した。化学の入試問題は、言語理解、化学知識の検索と利用、数学的推論の統合を必要とする、複雑な質問応答の課題であり、特に自然言語で書かれた問題文の意味解析のステップに言語処理に大きな課題がある。現在のシステムは骨組みだけ実装した段階で未対応の問題も多く、問題文から数値で解答が得られる問題は、センター試験過去 26 問のうち 4 問にとどまる。しかし、システムは現在未対応の問題にも適応可能な枠組みで設計されており、今後は 4 章で挙げたような問題に取り組んでいく。

参考文献

- [新井 12] 新井 紀子, 松崎 拓也, ロボットは東大に入れるか? : 国立情報学研究所「人工頭脳」プロジェクト, 人工知能学会誌 27(5), 463-469, 2012-09-01
- [Miyao and Kawazoe 13] Yusuke Miyao and Ai Kawazoe, University Entrance Examinations as a Benchmark Resource for NLP-based Problem Solving, Proceedings of IJCNLP 2013, Oct 2013
- [Barker 04] Ken Barker, Vinay K. Chaudhri, Shaw Yi Chaw, Peter E. Clark, James Fan, David Israel, Sunil Mishra, Bruce Porter, Pedro Romero, Dan Tecuci, Peter Yeh, A Question-Answering System for AP Chemistry: Assessing KR&R Technologies, In The Ninth International Conference on the Principles of Knowledge Representation and Reasoning (KR2004) 2004
- [工藤 02] 工藤 拓, 松本 裕治, チェンキングの段階的適用による日本語係り受け解析, 情報処理学会論文誌 43(6), 1834-1842, 2002