

# Question and Answering タスクの提案

福本 淳一†

加藤 恒昭‡

立命館大学 理工学部 情報学科†

東京大学大学院 総合文化研究科‡

fukumoto@cs.ritsumeai.ac.jp

kato@boz.c.u-tokyo.ac.jp

## あらまし

2002年開催予定のNTCIR3ワークショップに向けた新たなタスクとして Question and Answering Challenge (QAC) を提案している。QAC では分野独立の質問応答技術に関して、対話処理技術と情報抽出技術に力点を置いたタスクを定義する予定である。課題設定、評価方法、対象テキストについて現状を報告する。

## キーワード

質問応答, 情報抽出, 情報検索, 対話処理

## An Overview of Question and Answering Challenge (QAC) of the NTCIR3 Workshop

Jun'ichi FUKUMOTO† and Tsuneaki KATO††

Department of Computer Science, Ritsumeikan University†

Language and Information Science, University of Tokyo††

## Abstract

In this paper we will propose the question and answering task, called Question and Answering Challenge (QAC), and its first evaluation. This will be carried out as a sub-task of NTCIR3 Workshop scheduled in October 2002. In QAC, we aimed to encourage the development of practical Q&A systems in a general domain and focus on research of user interaction and information extraction. User interaction technology leads to actual interaction between computer and person. In actual Q&A between people, there will typically be several interactions in order to confirm the intention of the questions and so on. Information extraction that works in general domain is also an important technology in order to realize real Q&A system.

## Keywords

Q&A, Information Extraction, Information retrieval, user interaction

## 1 はじめに

我々は Q&A タスクの評価として、Question and Answering Challenge (QAC) の第 1 回評価会議 QAC-1 を 2002 年 10 月開催予定の NT-CIR3 ワークショップのサブタスクとして行うことを提案している [1].

Question and Answering(Q&A) は自然言語で記述された質問文に対して、大量の文書から適切な答えを得るタスクである。Q&A タスクは、長い間興味深い研究テーマとして考えられてきており、これまでに多くの研究がなされてきている。これら多くの Q&A システムはユーザの質問に対してシステム側とユーザ側のインタラクションの結果、答えが得られるものであるが、これらのインタラクションは限られた対象領域に関するものであった。現在の Q&A の研究は、対象領域を特定せず、一般分野の大量の情報から与えられた質問に対する答えを得ることがその目的となっている。

Q&A の技術は情報抽出、情報検索、自然言語インタフェースなどの自然言語処理の研究と関連がある。Q&A は、実際にユーザとの何度かのインタラクションによって答えを得る場合があるため、QAC においては、これらの研究分野の中で特にユーザとのインタラクションに注目する。また、与えられた質問から答えとなる要素を取り出すための技術として情報抽出技術に注目する。この情報抽出技術は、ある特定の分野で機能するのではなく、一般的な分野で機能するものを目指している。

Q&A は、テキストの自動要約技術とも関連している。質問がある特定の要素を答えとしているのではなく、「なぜ」や「どのようにして」などを求めている場合には、その答えはテキスト中の複数の文やテキスト全体の要約となると考えられるからである。また、Q&A の評価のためにはなぜそのような答えが得られたのかを説明として提示することが重要である。そのような回答の根拠となる情報としては答えの含まれているテキストの一部が考えられるがこの抽出もテキストの要約技術と関連していると考えられることができる。

以下では、QAC タスクと関連するいくつかの技術について、QAC との関わりについて述べた後、タスクの課題設定のポイントについて

述べる。最後に、現在予定しているタスクの概要について述べる。但し、このタスクは仮のものであり、今後の議論に基づいて改良していくための叩き台である。

## 2 QAC タスクの関連技術

質問応答 (Q&A) とは、大量の文書を背景に自然言語によって尋ねられた任意の質問に答えを与えるタスクであり、ドメイン依存でないこととともに、組織化されていない情報に依存していることが RDB に対する質問と異なる点である。Q&A をタスクとして捉えると、質問 Q に対する答え A の善し悪しの程度 (信頼度、蓋然性) が議論される技術であり、根拠を提示できることや答えのない質問に対処できる (誤るなら答えない) ことも重要な要素である。

### 2.1 対話処理技術

QAC タスクの最終的なゴールは、ユーザの質問に対して、システム側が大量の情報から適切な答えを返すことである。ユーザは関連する一連の項目について質問したい場合もあり、得られた答えに対して、その詳細情報や関連情報を追加的に質問することも考えられる。そのような一連の Q&A に対応するためには、対話処理技術が必要となる。また、ユーザの与えた質問内容が答えを絞り込むのに不十分であったり、あいまいである場合など、必要な情報を要求したり、あいまいな部分を解消するなど、システムからのインタラクションが必要になる場合もある (但し、このようなシステム側からの働きかけは今回の QAC では対象外としたいと考えている)。

### 2.2 情報検索・抽出技術

Q&A システムが与えられた質問に答えるためには、質問文中の情報によって大量のテキストからある程度の情報まで絞り込み、絞り込まれた範囲から答えとなる情報を取り出す必要がある。大量のテキストから情報の絞り込みを行うためには情報検索技術が不可欠である。さらに、テキスト中から必要な情報を取り出す技術として情報抽出技術があり、MUC[7] や IREX[8] などを中心に研究が進められてき

た。MUC では、英語の新聞記事を対象に固有名 (Named Entity)、参照関係 (Coreference)、属性情報 (Template Element)、要素間関係 (Template Relation)、シナリオ (Scenario Template) といった段階を設定し、それぞれ定義された抽出対象をテンプレートなどを用いて抽出を行ってきた。IREX においても、日本語の新聞記事を対象に固有名抽出をタスクとして設定してきた。

従来の情報抽出において必要な情報を得るためには、あらかじめ用意された抽出用のテンプレートを用いる必要がある、一般分野での任意の要素の抽出は困難なものであった。Q&A では、実際にどのような質問が行われるのかわかっていないため、答えとなる要素がどのようなものであるかが不明である。そこで、Q&A において答えとなる要素をテキストから抽出するために一般分野での情報抽出技術が重要な技術となると考えられる。

### 2.3 文書要約技術

Q&A においては、答えを得たときになぜそれが答えであるのかといった根拠となる情報を提示することも重要である。そのような情報としては、答えが得られたテキスト全体を示す方法やテキスト中の一部を提示する方法がある。さらに根拠となる情報がテキスト中の広範囲に分散している場合には、それらを要約して示すことも考えられる。したがって、文書要約技術も Q&A には重要な技術であると考えられる。また、「なぜ」や「どのようにして」などを求めている場合など、質問の種類によっては答えが語句レベルの情報ではなく、文レベルの情報が回答になる場合もある。文レベルの場合には、答えをテキスト中からそのまま抽出するだけでなく要約した形で答える場合も考えられ、この点からも文書要約技術は重要な技術である。

## 3 課題設定のポイント

Q&A タスクを設定する際には、質問文の設定、答え方、対象テキスト、評価方法などいくつかのポイントがある。

以下では、QAC における課題設定のいくつかのポイントについて述べる。但し、ここでは

今回のタスクのための具体的な課題設定だけでなく、それ以降のタスク設定についても視野に入れた上で述べる。

### 3.1 対象文書

質問の対象となるテキストとしては、どのようなものが適当であるか、一般に多く利用されている新聞記事データを利用することに加え、Web 上の文書を対象にすることも考えられる。Q&A の質問内容にも依存するが、質問が百科事典的な情報を尋ねている場合には別の知識源が必要になるがその場合にはどうすればよいかといった点が考えられる。現状では、著作権の関係で新聞記事データを対象として想定しているが、その場合に何年分の新聞記事を対象とするのがよいかも検討すべき項目である。

さらに、システムの対象を日本 (語) に閉じないためには、英語テキストも対象文書に含めた上で、それに対して日本語の Q&A にまで広げる可能性も存在する。多言語 Q&A システムなど、対象が何語でもよいといったものも考えられる。その際には、翻訳システムによるサポート方法なども研究対象となりえると考えられる。

### 3.2 システムの知識源

質問に答える上で、百科事典を持つてはいけない等、システムが利用する知識になんらかの制限を加えるべきかどうかを考える必要がある。もし、制限する場合、どのような制限が考えられるか。そもそも制限を規定することは可能か不可能かなども検討すべき項目である。

### 3.3 質問文のタイプ

質問文としては、5W1H 型の質問や yes/no 型のものが考えられる。Who, Where, When, What タイプの質問については、答えが名詞語句などの要素となり、Why, How などでは、答えが文レベルの表現となるものであると考えられる。また、最初の質問に関連した質問やさらに詳細な情報を尋ねるような質問の形式も考えられる。この際、後に続く質問では、最初の質問文から代名詞による言い換えや省略などの表現が多く用いられている。

また、実際のユーザとの対話を考えると、ユーザの与えた質問に対して、質問が曖昧であった場合や答えが非常に多く存在した場合など、システム側が何らかの問い返しを行い、ユーザがそれに答えたり、それに応じて質問表現を変更する場合もある。しかし、Q&A をタスクとして考えると、問い返し等、システムの応答に応じた動的な対応・課題設定が必要となり、これをタスクとして設定するのは非常に難しいため、QAC においては対象外とする。

質問については、すべて、正解がただひとつ存在するとは限らず、正解が対象テキスト中に存在しないことや、複数存在することがありえる。このような条件を想定した場合、答えが必ずひとつだけ存在するという前提でシステムを作成するのは全く異なった条件でのシステム作成が必要になると考えられる。

### 3.4 質問文の作成

質問文の作成について、現実的(実際になされそうな質問)で、かつ作成の手間が現実的なものとなる方法論はどんなものか、どのように進めればよいかについても検討を行う必要がある。Q&A システムの評価のための体系的な質問文の分類についても質問文作成のためだけでなく、システムの評価のためにも重要な技術であると考えられる。

### 3.5 評価方法

Q&A の評価は、質問に対して正しく答えが得られたかがその対象となる。また、答えがどこから得られたのかを示す根拠を示す情報についても答えが正しいのかを判断するための重要な情報であるため評価対象とすることも考えられる。

TREC の QA タスク [2][3][4] や NTT CS 研による報告 [5] では、上位 5 個までの答えを返し、正解の順位の逆数と得点とする評価が行われている。例えば上位のいくつかを評価対象にするのか、また、正解のみを評価対象とするかなどの評価方法がある。根拠情報についても評価に含めるか、含めるとした場合どのような形で含めていくかも評価の重要な観点である。

## 3.6 評価後の言語資源

Q&A について、評価の後に資産として残すものにはどのようなものがあるのかについても検討を行う必要がある。例えば、要約であれば、要約結果のデータを言語資源として残すことも評価のための大きな目標になっている。Q&A の場合には、最終的に残る言語資源としては、質問と答えのセット、および、答えの根拠情報が言語資源として残るものであると考えられる。

## 4 設定するタスクについて

現在想定している QAC タスクの内容について述べる。タスクの設定のポイントは以下のとおりである。

- 情報抽出技術、特に固有名抽出技術を土台とした Q&A システムが研究の背景にある。しかしながら、人名、組織名、地名などの固有名となっている要素だけでなく、それ以外についても、例えば、役職名などの属性的な情報を答えとするような質問を設定する。また、固有名だけでなく、要素の属性情報 (TE) や要素間の関係 (TR) の扱いも要素技術として想定している。
- 質問文を単独のものではなく枝問的にすることで、質問文どうしもしくは答えとそれに続く質問文との参照関係を解釈することで後続の質問の解釈が可能になるような質問文を想定している。
- 答えのないものや答えが複数ある質問文を想定している。これは、答えが対象テキスト中に存在しないような質問にどのようにして答えるべきか、また、答えが複数ある場合の答え方の研究を目指したものである。

### 4.1 対象テキストを何にするか。

NTCIR の別のサブタスクである TSC2(要約タスク)とも内容を合わせるということから、毎日新聞の 98,99 年の 2 年分が候補として考えられる。QAC のタスクとしては、これら 2 年

分が対象として適切であるかどうか、どちらか1年分とするかを定める必要がある。

## 4.2 答えの範囲

今回は、質問への答の候補として、固有名、その他の名詞、名詞句、数値表現を想定している。答えの範囲として例えば名詞句という定義を行った場合、ここでいう名詞句とは何か、形態素解析の返してくる値としての名詞句であるのか、数値表現は含むのかといった議論が起こってくる可能性がある。これを避けるために今回は目安として固有表現のような名詞、名詞句、数値表現など(いわゆる体言の句、ただし小説や映画のタイトル等を含む)を答えとして想定しているというレベルの説明ということで、明確な定義は行わない方針である。

## 4.3 質問文の設定

以上の答えの範囲を想定して、代名詞や省略などで関連付けられた、連続した質問を設定する。質問とそれに関連した質問を1つから3つ程度与える。質問については、すべて、正解がただひとつ存在することを仮定しない。つまり、正解が知識源中に存在しないことや、複数存在することがありえるという仮定である。ただし、これらはサブタスクとする。サブタスクの具体的な仕様案については後述する。最も基本的なタスクは前年の TREC 相当のものとする。

また、質問文の作成について、どのように進めればよいか。現実的(実際になされそうな質問)で、かつ作成の手間が現実的なものとなる方法論はどんなものかについて検討する必要がある。

## 4.4 質問の数

1タスクあたり、100程度の質問数を予定している。

## 4.5 根拠情報

システムが答えとして返してきたものが正しいかどうかを判断するための情報として根拠情報が必要であると考えている。必要な場合、根拠情報としては、テキスト中から抽出されたあ

る範囲のテキストとする。例えば、答えの根拠となる情報を含む1パラグラフや100文字以内の複数文などがある。但し、根拠情報についてどのような情報が適切であるのかについて、今後議論する必要がある。

## 4.6 タスクの例

### [タスク 1]

システムは、与えられた質問に対して、その回答と考えられるものひとつを優先順位をつけて5つ返す。評価は、最も高い優先順位を与えられた正解(もしくは複数正解のうちのひとつ)に基づいて行われる。例えば、順位の逆数(1位ならば1, 2位ならば1/2等)による得点付け等。正解(のいずれかひとつ)をなるべく高い順位で解答したものが高い評価を得る。すべての正解を網羅する必要はない。誤った回答に対してペナルティは与えない(従って、正解の存在しない質問は得点に影響しないので、ないのと同じ)。基本的には TREC QA Task と同じ課題設定となる。

例えば、正解が、山田と鈴木のみであるような質問に対して、システムは、例えば、第一候補、佐藤、第二候補、鈴木、第三候補、田中、第四候補、山田、第五候補、山本、の様に解答する。この場合、第二候補に正解のひとつである鈴木があげられているので、それに基づいて得点が計算される(例えば、2の逆数の1/2点等)。第四候補にも正解があげられているが、これはより高い優先順位の正解があるので、あってもなくても得点は同じである。

### [タスク 2]

システムは、与えられた質問に対して、その回答と考えられるもののリストをただひとつ返す。評価はそのリストに含まれるものと正解との差分に基づいて行われる。正解であるものを含んでいなかったり、正解でないものを含んでいたりする場合は、いずれも減点となる(計算方式の詳細は未定)。従って、正解のない質問に対しては、空リストを返すと得点となるが、それ以外のものを返すと減点の対象となる。

例えば、正解が、山田と鈴木のみであるような質問に対して、システムは、例えば、(山

田, 鈴木), (鈴木), (佐藤, 山田, 鈴木) 等のいずれかを回答する。(山田, 鈴木) の場合は, 正解のすべてを網羅し, 余計なものを含んでいないので, 満点. それ以外のふたつの例は正解を網羅していなかったり, 余計なものを含んでいるので減点の対象となる.

タスク 1 向けにシステムを設計した場合でも, 候補を並べるためになんらかの尤度計算をしているのであれば, その尤度に閾値を設定して, その閾値以上の候補をリストして回答することでタスク 2 にも参加できる.

タスク 1 とタスク 2 は同じ質問セットを使うことを考えている.

### [タスク 3]

タスク 2 の形式で回答するシステムに対して, 複数の連続した質問 (枝間) に回答させる. 連続した質問の最初 (先頭) のものは, タスク 2 で用いられる質問と同じ (この部分をタスク 2,3 で共有するかは要検討) とする. それに続いて, 先行する質問やその回答を参照する質問が行われる. 評価はそれぞれの質問について, タスク 2 と同様の形式で行う.

根拠情報はタスク 1 の場合, 各候補についてひとつ, タスク 2 の場合は, リストの各要素についてひとつ, ただし, ある部分がリスト内の複数要素のサポートとなっている場合は共有して構わない. それぞれの根拠情報は 1 パラグラフもしくは 100 文字程度. 提出することは義務付けるが, 採点の対称とするかは今後検討していく.

## 4.7 スケジュール

現在のところ想定している評価スケジュールを以下に示す. ただし, これはあくまでも予定であって, ワークショップ以外のスケジュールは仮のものである.

dry run	2001.12
formal run	2002.3
NTCIR3 workshop	2002.10

## 5 おわりに

本稿では, 2002 年開催予定の NTCIR3 ワークショップに向けた新たなタスクとして提案を

行っている Question and Answering Challenge (QAC) について, 関連する技術の観点から評価の目的について述べるとともに, タスクの設定について, 課題設定, 評価方法, 対象テキストなどいくつかの項目について述べた. 本評価プロジェクトの CFP は 2001 年 7 月を予定しており, 多くの団体に参加してもらえることを希望している. また, 本稿で述べたタスクについては不十分な点も多く, 今後, メーリングリスト (qac@nii.ac.jp, qac-j@nii.ac.jp<sup>1</sup>) や検討会を通じてよりよいタスクにしていければと考えている.

## 謝辞

本稿をまとめるにあたって, 多くの貴重な助言を頂きました QAC タスク検討会に参加された皆様, および, QAC-ML で議論に参加された皆様に感謝いたします. また, このような評価の機会を与えてくださいました, 国立情報学研究所の神門先生に感謝いたします.

## 参考文献

- [1] Fukumoto, J. and Kato, T. An Overview of Question and Answering Challenge (QAC) of the next NTCIR Workshop, Proceedings of the Second NTCIR Workshop Meeting, National Institute of Informatics, 2001.
- [2] Burger, J., Cardie, C. et al. Issues, Tasks and Program Structures to Roadmap Research in Question & Answering (Q&A) NIST DUC Vision and Roadmap Documents <http://www-nlpir.nist.gov/projects/duc/roadmapping.html>, 2001.
- [3] Voorhees, E.M. and Harman, D.K. (eds.) Proceedings of the Eighth Text REtrieval Conference (TREC-8)

<sup>1</sup>ML に参加を希望される方は, majordomo@nii.ac.jp へ subscribe qac-j@nii.ac.jp ご自身のメールアドレス (日本語用) subscribe qac@nii.ac.jp ご自身のメールアドレス (英語用) とメールメッセージの本体に書いたメールを送ってください. 自動登録されます. 自動登録の方法につきましては, 下記のアドレスを参照してください. qac-j@nii.ac.jp への登録者は自動的に qac@nii.ac.jp にも登録されます.

- [http://trec.nist.gov/pubs/trec8/t8\\_proceedings.html](http://trec.nist.gov/pubs/trec8/t8_proceedings.html), 2000.
- [4] QA Track Overview TREC9,  
<http://trec.nist.gov/presentations/TREC9/qa/index.htm>, 2000.
- [5] 佐々木, 磯崎, 平, 廣田, 賀沢, 平尾, 中島, 加藤, 質問応答システムの比較と評価, 信学技報, NLC 2000-24, pp.17-24, 2000.
- [6] Voorhees, E.M. and Tice, D.M. Building a Question Answering Test Collection, Proceedings of SIGIR2000, pp. 200-207, 2000.
- [7] Proceedings of 7th Message Understanding Conference (MUC-7), DARPA, 1998.
- [8] Information Retrieval and Extraction Exercise (IREX), <http://cs.nyu.edu/cs/projects/proteus/irex/>, 1999.