社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2008－ＦＩ－ 89 (7)
2008－ＮＬ－183 (7)
2008／1／21

NTCIRにおける質問応答技術の評価と今後の展望

森 辰則[†] 福本 淳一[††] 加藤 恒昭[†††] 桝井 文人[‡]
佐々木裕[‡‡] Hsin-Hsi Chen[‡‡‡] Kuang-hua Chen[‡‡‡] Chuan-Jie Lin[*]
三田村 照子[**] Eric Nyberg[**] 神門 典子[***]
[†]横浜国立大学, [††]立命館大学, [†††]東京大学, [‡]三重大学,
[‡‡]The University of Manchester, [‡‡‡]National Taiwan University, [*]National Taiwan Ocean University,
[**]Carnegie Mellon University, [***]国立情報学研究所
E-mail: mori@forest.eis.ynu.ac.jp

本稿では，国立情報学研究所が主催する評価型ワークショップであるNTCIRにおいて実施された質問応答技術の評価について解説をするとともに，現在進行中の言語横断情報アクセス技術に関する新しい評価の枠組であるACLIA (Advanced Cross-lingual Information Access)について述べ，情報アクセス技術の評価に関する今後を展望する．

Evaluation of Question-answering-related Technologies in NTCIR

Tatsunori Mori[†] Jun'ichi Fukumoto[††] Tsuneaki Kato[†††] Fumito Masui[‡]
Yutaka Sasaki[‡‡] Hsin-Hsi Chen[‡‡‡] Kuang-hua Chen[‡‡‡] Chuan-Jie Lin[*]
Teruko Mitamura[**] Eric Nyberg[**] Noriko Kando[***]
[†]Yokohama National University, [††]Ritsumeikan University, [†††]The University of Tokyo, [‡]Mie University,
[‡‡]The University of Manchester, [‡‡‡]National Taiwan University, [*]National Taiwan Ocean University,
[**]Carnegie Mellon University, [***]The National Institute of Informatics
E-mail: mori@forest.eis.ynu.ac.jp

In this paper, we described the previous evaluation about question-answering techniques in NTCIR, which is a series of evaluation workshops organized by the National Institute of Informatics, Japan. We also explained a new scheme of evaluation for cross-lingual information access in NTCIR, called ACLIA (Advanced Cross-lingual Information Access), with some future prospects.

1 はじめに

テキストを対象にした情報アクセス技術において，重要なものの一つが質問応答技術である．質問応答技術とは，利用者が明確な情報要求を持っており，それを自ら自然言語の質問文の形で表現できる時に，その質問文の答となる表現を大量の文書コレクションの中から根拠情報とともに抽出するための技術である．

他の情報アクセス技術と同様に，質問応答はTREC[2]やCLEF[3]といった評価型ワークショップの中でその技術が磨かれてきている．特に，アジア圏の言語を対象とする研究者に対しては，国立情報学研究所 (NII)が主催する評価型ワークショップNTCIR (NII Test Collection for IR Systems)が国際研究交流のフォーラムを提供してきた[1]．本稿では，NTCIRにおいて実施されてきた質問応答関連の技術に対する評価と，その過程で登場した各種技術について解説をするとともに，現在進行中の言語横断情報アクセス技術に関する新しい評価の枠組であるACLIA (Advanced Cross-lingual Information Access)について説明し，情報アクセス技術の評価に関する今後を展望する．

2 質問応答処理技術

質問応答とは，利用者から自然言語の質問文を受け取り，それに対する解答を情報源となる文書コレクションからみつけ，利用者に提示する技術である．過去に人工知能研究では「質問応答」として，限定した領域において組織化された知識や規則を用意し，それらに基づいて推論を行ない，利用者との対話を行なう仕組みが研究されていた．これと区別するために，我々が対象とするような，大量文書を情報源として用い，特定の領域に関する知識の整備を必要としない質問応答は，「領域非依存質問応答 (open domain question answering)」と呼ばれることもある．

質問応答技術は，いくつかの観点によって複数の種類に分類することができる．

- 回答の種類

  factoid 型 名称 (人，組織，製品等の名前) や数に纏わる表現 (金額，大きさ，日付等) など短い表現 (名詞句) が回答となる．

  non-factoid 型 descriptive型, complex型と呼ばれることもある． 定義，理由，方法，関

係など，名詞句よりも大きな単位 (節，文，段落) が回答となる．複数箇所から得られた情報を統合・要約することもある．

- 言語 (質問，情報源となる文書コレクション)

  単言語 質問と文書コレクションの言語が一致．

  言語横断 両者の言語が異なる．

- 情報源となる文書コレクション

  手元の文書コーパス 電子化された新聞記事など手元で加工が可能な文書集合．

  World Wide Web ネットワークで公開されている文書．

- 文脈依存性/対話

  一問一答型 それぞれの質問文中で十分な文脈が与えられ，各質問に独立に答える．

  文脈依存/対話型 以前の質問応答を文脈として，現在の質問に答える．

- 回答の仕方

  優先順位型 解候補に優先順位をつけ，上位の解候補から順番に指定した数だけ出力．

  リスト型 正解と思しき解候補を過不足なく列挙．

処理の観点からいうと，質問応答システムは図 1 に示すとおり，複数のモジュールから構成される複合システムであり，主に「質問文解析」，「文書/パッセージ検索」，「解候補の抽出」，「解の選択」からなる．

質問文解析部では，利用者が入力した質問文を解析し，質問文の型，質問の焦点，質問文中のキーワードの列などを抽出する．文書検索部では，質問文中のキーワードの列を受けとり，関連文書の検索を行ない処理対象の文書を絞り込む．また，文書よりも小さい単位であるパッセージを検索対象とすることもある．解候補の抽出部では，文書中の言語表現と質問文の間の類似度や，質問文の型と解候補の種類の間の一貫性などをスコアにし，尤もらしい解候補を抽出する．解の選択部では，抽出された解候補に対し，推論や投票などを行ない，解の妥当性判定を行なう．

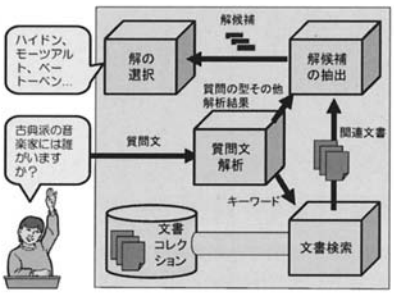

図 1: 質問応答システムの構成例

factoid 型質問応答においては，短い表現，特に名称が解候補になるので，対象文書に固有表現抽出器を適用し，抽出された固有表現を解候補とすることが多い．

一方で，non-factoid 型の質問応答においては，解候補が文の系列など長い表現となる．質問文の型により規定される解候補の記述様式の適切さや，解候補と質問文の関連性等に基づき解候補を決めることが多い．

言語横断型の場合においては，質問文と文書コレクションの言語が異なるため，いずれかの段階で質問もしくは文書の翻訳が行なわれる．その際に，対訳辞書や機械翻訳システム等が用いられる．

また，以前の質問応答の結果を踏まえて，利用者が次々と質問をする場合を想定すると，以前の質問や回答を参照した質問が入力されることが予想される．このような対話型の質問応答システムにおいては，質問文中の参照表現を解析する仕組みが必要とされる．

3 NTCIRにおける質問応答技術評価

NTCIR における質問応答技術評価の経緯を図 2，3 に示す．NTCIR における質問応答技術評価は，他の評価と同様に，共通の評価用テストコレクションを用いて行なわれてきている．評価用のテストコレクションは通常以下の各データから構成されている．

1. 質問文のコレクション
2. 情報源となる文書コレクション
3. 各質問毎の正解の情報

正解情報は，質問文の正解である文字列とその根拠を与える文書の識別子の組の集合である．factoid 型質問応答のように文書中の表現をそのまま抽出したものが解候補となる場合には，各参加システムが抽出した解候補を人間の判定者が調べ，正解となる抽出表現をそのまま保持する．このように作成された正解情報と別途定義された評価尺度に基づき，システムの出力を自動評価する自動採点ツールが提供されることが多い．

一方で，non-factiod 型質問応答の場合には，回答が長い記述となり，問題設定によっては複数の文書断片を多少加工した上で結合することもあり得るので，システム出力に対して判定者 (人間) が統一された基準において判定を行なう必要がある．

あるシステムにより，テストコレクションの質問文をすべて処理をして回答を得る過程をラン (run) という．タスク設定によっては，公式ラン (formal run) 以外に，公式ランの前に行なう試行であるドライラン (dry run) などが行なわれることがある．

ここで，上記の構成による評価用コレクションは，システム全体の入力 (質問文) と出力 (回答) のみを規定しているので，質問応答過程全体を通じた精度の評価はできるが，質問応答処理を構成する各モジュールの評価はできないことに注意されたい．そのようなモジュール毎の評価を行なうためには，途中結果に対する評価を別途行なわなければならない．

さてNTCIRにおける質問応答に関連するタスクとしては，以下のものがある．いずれも，数年分の新聞記事を文書コレクションとしている．

- QAC (Question Answering Challenge): NTCIR-3〜NTCIR-6で実施される．
- CLQA (Cross-lingual Question Answering): NTCIR-5〜NTCIR-6で実施される．
- ACLIA (Advanced Cross-lingual Information Access): NTCIR-7で実施中．

QACは日本語を対象とした単言語のタスクであり，質問文の型としてはfactoid型，non-factiod型，回答の仕方としては優先順位型，リスト型が対象となった．更に対話型タスクとして，Information Access Dialogue (IAD) タスクも行なわれた．詳細は節4で述べる．

また，CLQAは言語横断のタスクであり，質問文の型としてはfactoid型，回答の仕方としては優先順位型が対象となった．詳細は節5で述べる．

一方，ACLIAはNTCIR-7におけるクラスタと呼ばれるタスクの集合体の一つである．ACLIAではQACとCLQAで行なわれていたタスクを統合/発展させた新しいタスクCCLQA (Complex Cross-Lingual Question Answering) が実施されている．ACLIAには，これ以外に，IR for QAがタスクとして含まれる．詳細は節6で述べる．

4 NTCIR QAC

4.1 NTCIR-3 QAC

NTCIR-3 QACはNTCIRにおいて最初に行なわれた質問応答タスクであり，factoid型質問応答の評価が3つのサブタスクにおいて行なわれた [5]．いずれのサブタスクでも，文書コレクションとして，1998年，1999年の毎日新聞記事が用いられた．他の情報源を処理に利用してもよいが，回答の根拠として，新聞記事による裏付けが必要であった．質問に対する解は文書コレクション中の表現を抽出したものであり，名称(人，組織，製品，地域等の名前)や数量表現(金額，大きさ，日付等)を表す名詞句である．

サブタスク1は，各質問に対して優先順位型で5つの解候補を求めるタスクであった．質問コレクションは200問から構成されている．評価の尺度は，最上位の正解の順位の逆数であるReciprocal Rankについて，全質問に亙って平均値を求めたMRR (Mean Reciprocal Rank) が用いられた．この尺度は，[0,1]の範囲の値を持ち，1に近いほど求解精度が良い事を示す．

サブタスク2は，各質問に対してリスト型で解候補の集合を求めるタスクであった．質問コレクションとしては，サブタスク1と同じものが用いられた．評価の尺度には，平均修正F値 (Mean Modified F-Measure, MMF) が用いられた．修正F値は，通常のF値とほぼ同じであるが，正解が無い質問についての扱いが異なる．すなわち，そのような質問に対して，空リストをシステムが出力したときにはF値として1を与え，それ以外は0を与える．

サブタスク3は，サブタスク2と同じくであるが，各質問に関連質問が1問後続した．この後続質問に対するリスト型質問応答が評価の対象となった．後続質問には先行する質問やその回答を参照する表現があるので文脈依存型タスクであった．質問は40問あった．

各サブタスクについて，質問の例や，参加者が用いた各種技術，求解精度については，図2におけるNTCIR-3の列を，また，質問コレクションや各参加者の評価結果は，[14] を参照されたい．

4.2 NTCIR-4 QAC

NTCIR-4 QACは，NTCIR-3 QACとほぼ同様のタスク設定であったが，以下の各点が改訂された [6, 8]．なお，公式ランで利用された質問コレクションは，サブタスク1においては，200問が新たに作成され，サブタスク2においては，以下に述べるようにサブタスク1とは独立に200問が準備され，サブタスク3においては，36の質問系列，計251問が用意された．

情報源 文書が追加され，1998年，1999年の読売新聞記事も情報源として利用された．

サブタスク2の質問コレクション サブタスク1とは異なる質問コレクションを用いた．NTCIR-3 QACにおいては，各問の正解の数が元々少なかったために，上位1位の解を出力するという単純な戦略でも高い平均F値となってしまったこともあり，複数の解がある問題を別途用意した．

サブタスク3の後続質問 後続質問文の数が1から平均5.92へと増えた．また，後続質問における主題の扱われ方をgathering型，browsing型という2種類に大別し，これらが混在した質問系列のコレクションが準備された．gathering型は，最初の質問文で導入された一つの主題が後続する質問でも受け継がれる質問系列の型である．一方で，Browsing型は，対話が進むとともに主題が変化していく質問系列の型である．

サブタスク3では，後続質問が増えたことにより，利用者がシステムとの対話により複合的な情報アクセスを行なう状況を模擬するように，後続質問が設定されている．そのため，Information Access Dialogue (IAD) タスクと称されている．

各サブタスクについて，質問の例や，参加者が用いた各種の技術，求解精度については，図2におけるNTCIR-4の列を，また，質問コレクションや各参加者の評価結果は，[15] を参照されたい．

4.3 NTCIR-5 QAC

NTCIR-5 QACでは，それまでのサブタスク1，2は継続されず，サブタスク3であったIADにタスク

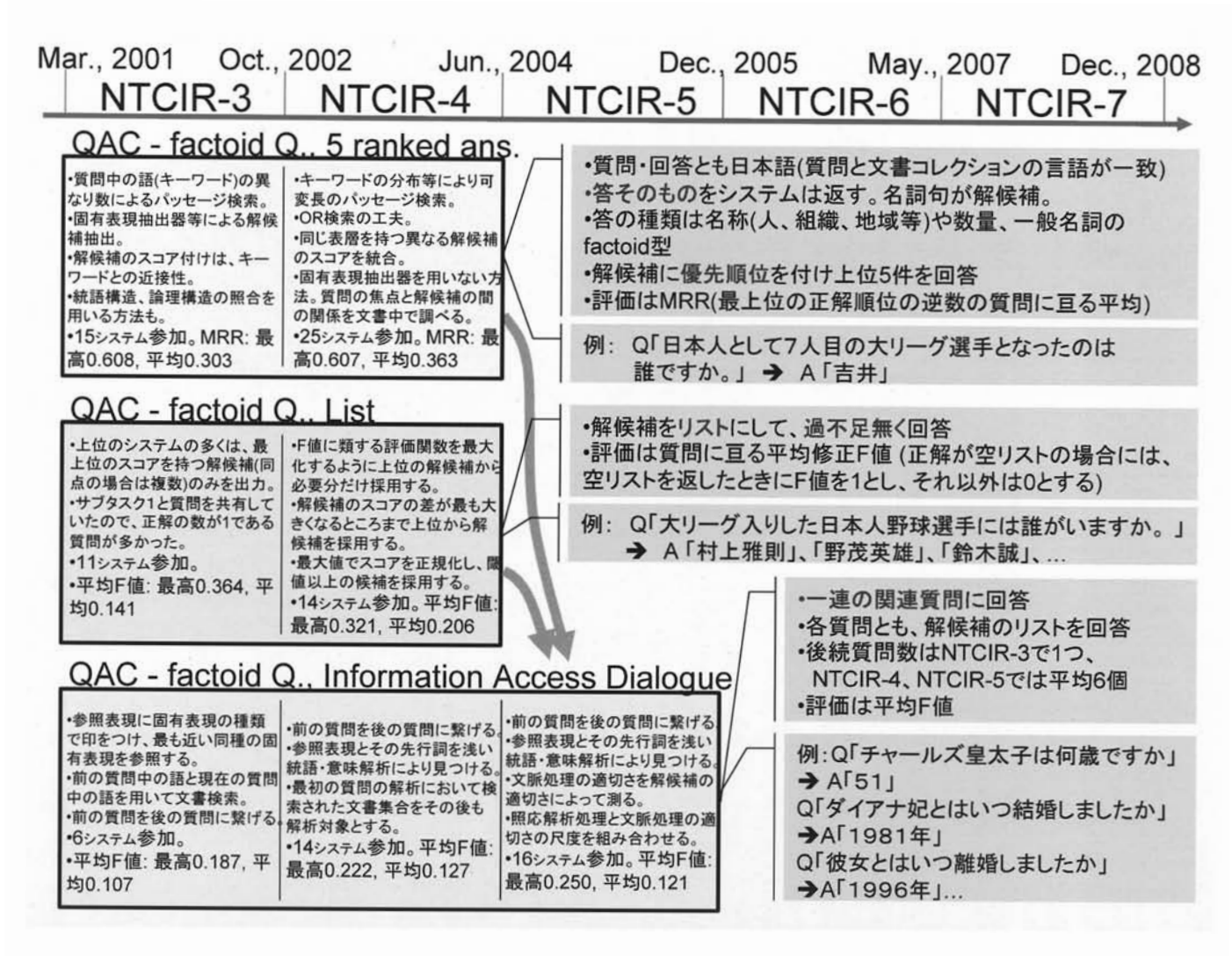

図 2: NTCIR における質問応答関連タスクの変遷 (1)

図 3: NTCIR における質問応答関連タスクの変遷 (2)

が集約された [9]．公式ラン用に 50 の質問系列，合計 360 質問が準備された．質問系列の内訳は，gathering 型が 35, browsing 型が 15 であり，1 系列の平均質問数は 7.2 問であった．また，平均正解数は 1.98 であった．

以前のタスクと比較して，ここでの IAD タスクは以下の各点が改訂された．なお，質問の例や，参加者が用いた各種技術，求解精度については，図 2 における NTCIR-5 の列を，また，質問コレクションや各参加者の評価結果は，[16] を参照されたい．

情報源 情報源となる文書コレクションが，2000 年，2001 年の毎日新聞，読売新聞の記事になった．

解の範囲の再定義 NTCIR-3 QAC で定義された範囲に加えて，付加的な表現を伴う数量表現 (例「長さ 30cm」)，範囲表現 (例「10〜12パーセント」)，概数 (例「約 100 人」)，の他，名詞句の範囲で答えられる，定義，特徴，理由，状況などあるイベントを表す表現も解の範囲となった．

評価の精密化 正解を過不足無く列挙する方法は，複数存在することがある．その列挙方法の各々を correct answer set (CAS) と呼び，それらに基づいて評価を行なっている．

Wizard of Oz 方式による質問作成 あるトピックについて，専門家が質問応答システムを模擬する形で，利用者を模擬する被験者からの質問に回答する．その時の質問の系列を利用する．

4.4 NTCIR-6 QAC

NTCIR-6 QAC では，factoid 型から non-factoid 型へと回答の種類が変更された [7]．公式ランでは，定義，理由，方法等が回答となる non-factoid 型質問を含む合計 100 問の質問が用意された．システムはリスト型で出力を行なう．

non-factiond 型の質問応答では，節や文，段落といった比較的長い答が期待されるとともに，複数の断片をつなぎ合わせて一つの回答を得ることが必要な場合もあり得る．よって，文書からの短い表現の抽出であった factoid 型質問応答で利用された評価方法は使えない．そのため，NTCIR-6 QAC-4 では，質問応答トラック以外に評価トラックが用意され，評価方法の提案もタスクの一部となった．

さて，タスクオーガナイザ側の評価は以下のように行なわれた．まず，参加各団体に依頼をし，公式ラン用の 100 問について，それぞれ人手で正解を作成して頂き，これを元にして正解情報を作成した．この正解情報を基に，システムの出力した解候補の評価が人手によりなされた．参加 14 システムが出力した解候補の総計は 14050 であったが，これらすべてを人手で評価することは時間的な制約のために難しく，各団体に解候補の絞り込みを依頼した．返答が無い団体の出力については，リストから 4 件を採用した．このようにして得られた総計 3750 の解候補を人手で評価した．判定者は 2 名であったが，質問毎にみると，いずれかの判定者 1 人が判断をしている．評価は以下の 4 基準により行なわれた．Level A,B,C の回答が正解に関する何らかの情報を含んでいる回答である．

レベル A 正解とほぼ等しい内容

レベル B 正解に加えて他の情報を含む．そこでの主たる内容は対応する正解の記述ではない．

レベル C 正解の部分ではあるが，全体ではない．

レベル D 正解の内容を含まない．

なお，今回の評価においては，システムの性能を一つの指標で示すような公式な評価尺度は定義されていない．上記評価結果を基に各参加者が自分のシステムを分析した．ただし，参考情報として，「そのシステムが正解を見つけることができた質問の割合」や「全出力における特定レベルの解候補の割合」などいくつかの観点でのシステムの比較検討結果が参加者に提供され，タスク概要論文で示された [7]．

一方，評価トラックにはオーガナイザの一人が主宰する 1 団体が参加した．自動要約の自動評価のために提案されている Basic Element(BE) を用いた手法を質問応答の評価に利用することが提案されている [4]．

なお，質問の例や，参加者が用いた各種技術，求解精度については，図 3 における NTCIR-6 の列の QAC の項目を，また，質問コレクションや各参加者の評価結果は，[12] を参照されたい．

5 NTCIR CLQA

NTCIR CLQA は factoid 型かつ優先順位型の言語横断のタスクであり，NTCIR-5,NTCIR-6 において実施された．言語横断の種類としては，NTCIR-5 では，JE, EJ, CE, CC, EC の組み合わせが実施され，NTCIR-6 では，これらに加えて，単言語型の JJ, EE の組み合わせも CLQA の枠組で行なわれた．ここで，言語横断の種類を表す “XY” という記号は，質問と文書コレクションがそれぞれ，言語 X，Y で記されている事を表す．J は日本語，C は中国語 (繁体字)，E は英語である．

質問コレクション中の各問は 0 もしくは 1 つの正解をもつ．解は名称 (人，組織，製品，地域の名) や数量表現等の固有表現に限定されている．数量表現には概数表現も許している．

公式ランでは各問に対して文書コレクションの言語で記述された解候補を 1 つだけ出力することが各システムに求められた．実際の言語横断質問応答の使用状況では，回答を質問側の言語に翻訳することが期待されるが，これは非公式ランにおいて推奨されている回答方法である．なお，非公式ランでは，各システムは各質問に対して最大 5 個の回答を順位をつけて出力することが求められた．

システムが出力した回答の各々について，正解情報に基づき，次の評価が付与されシステム評価の原情報となる: R(正解であり，根拠文書も正しい)，U(正解であ

るが，根拠文書は違っている)，W(間違い)．評価尺度は，Accuracy，MRR，Top5が用いられた．Accuracyは公式ランに対する尺度であり，全質問に対してシステムが正解を出力した質問の割合である．MRRならびにTop5は非公式ランに対応する尺度である．Top5は，5つの解候補に正解がはいっていた質問の割合である．

以下では紙面の関係で日本語に関連するサブタスクに限って説明をする．

5.1 NTCIR-5 CLQA EJ,JEサブタスク

NTCIR-5 CLQA[17]における，EJならびにJEサブタスクでは，文書コレクションとして，日本語は2000年，2001年の読売新聞，英語は同じ期間のDaily Yomiuriが用いられた．

質問コレクションの構築方法は次のとおりである．まず，あるトピックについて英日で対応する記事の組を選択し，英語の新聞を読んで英語の質問を作成する．次に，英語質問を作成する際に参照した英語記事と対になっている日本語の記事を参照し，それに基づいて対応する質問を作成する．このため，英日の質問文対は記事に依存する形で対訳になっている．

公式ランの前にドライランの替わりとして配布されたサンプルデータにおいては，300の質問と回答の組があった．一方，公式ランの質問コレクションは200問から構成されていた．

なお，質問の例や，参加者が用いた各種技術，求解精度については，図3におけるNTCIR-5の列を，また，質問コレクションや各参加者の評価結果は，[10]を参照されたい．

5.2 NTCIR-6 CLQA EJ,JEサブタスク

NTCIR-6 CLQA[18]における，EJならびにEJサブタスクでは，文書コレクションとして，日本語は1998年，1999年の毎日新聞，英語は同時期の，Taiwan News, China Times English News, Mainichi Daily News, Korea Times, Hong Kong Standardが用いられた．

NTCIR-5 CLQAにおいて採用された質問作成手法では，質問が情報源となる新聞記事の記述様式に過度に依存してしまうという問題があった．そこで，NTCIR-6 CLQAにおいては，対訳になっている新聞記事を利用せず，公式ランで使用される日本語文書コレクションとは別の新聞記事を参照しながら質問作成をし，そのうちの200問を公式ランで使用している．

なお，参加者が用いた各種技術，求解精度については，図3におけるNTCIR-6の列のCLQAの項目を，また，質問コレクションや各参加者の評価結果は，[11]を参照されたい．

6 ACLIAにおける評価の概要と展望

ACLIAはNTCIR-7におけるクラスタと呼ばれるタスクの集合体の一つである．公式ランは2008年6月に予定されている．ACLIAの目指す最終的な目標は，任意の型の質問に対し，多言語の情報源から解候補を抽出し，利用者が指定した言語で回答するという，究極の言語横断情報アクセスである．factoid型質問のような簡単な情報要求だけではなく，複数の文書から質問の回答を統合し，要約することが必要となる，より複雑な情報要求も扱うことが想定されている．

この目標に向けて，ACLIAでは，QACとCLQAで行なわれていたタスクを統合/発展させた新しいタスクCCLQA (Complex Cross-Lingual Question Answering)が現在進行中である．CCLQAは言語横断のnon-factoid型質問応答タスクである．従来のNTCIR, TREC, CLEFではこの設定における質問応答は扱われてきておらず，NTCIR-7において初めて評価されるものである．また，ACLIAには，CCLQA以外にCLIR for QAがタスクとして含まれる．言語横断の種類としては，CLQAとしては，EJ，JJ，EC，CCがあり，CLIRとしてはEJ，ECがある．ただしCLQAにおける'C'は繁体字もしくは簡体字の中国語であり，CLIRにおける'C'は簡体字である．

取り扱う質問応答の違い以外に，ACLIAではシステムのモジュール毎の評価を丁寧に行なおうとしている．言語横断質問応答システムは，通常，様々なモジュールから構成される複合的なシステムである．従来，システム全体の入力(質問)と出力(回答)の間の関係を調べ，システムの総合評価をしていたが，システムの改善を目指して詳細な評価を行なうためにはモジュール毎の評価が必要である．そのためACLIAでは，図4に示すとおり，典型的なシステム構成によりいくつかのモジュールを設定し，モジュール間のデータの受渡しをXMLによって定義することにより，これらに従って構成されたシステムを，モジュール毎に評価をすることが計画されている．特に，文書検索(CLIR)はCLQAにおいて本質的な部分を担うので，質問応答の観点から，モジュール単体の性能，ならびに，これが部品として埋め込まれた質問応答システム全体の性能について，評価がなされる予定である．

質問としては，関連イベントのリスト，人物情報や定義，エンティティ間の関係を問うものが扱われる予定である．なお，質問の例は，図3におけるNTCIR-7の列を参照されたい．また，ACLIAのためのWikiページが用意されているので，各種情報はここから取得可能である[13]．

7 おわりに

本稿では，NTCIRにおいて実施されてきた質問応答関連の技術に対する評価と，その過程で登場した各種技術について解説した．また，現在進行中の新しい評価の枠組であるACLIAについて説明をし，情報ア

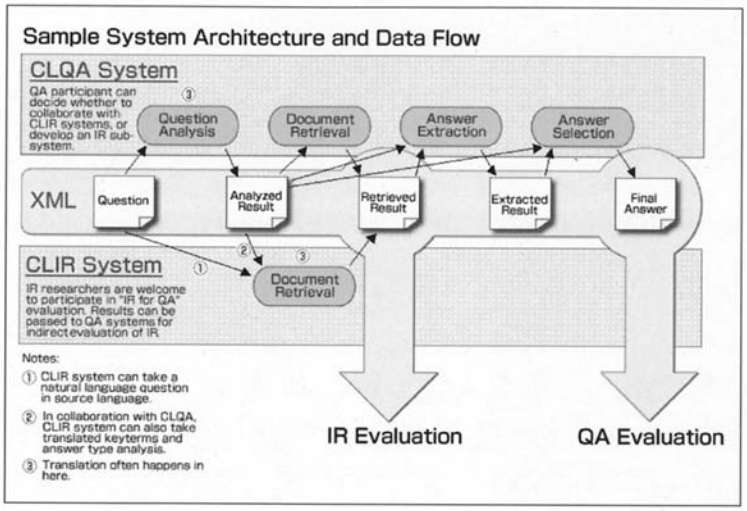

図 4: ACLIA: Advanced Cross-lingual Information Access

クセス技術の評価に関する展望を述べた.

NTCIR のような評価型ワークショップの成否は，広範な方々の積極的な御参加に掛かっている．少しでも興味をお持ちであれば，ぜひ，ACLIA をはじめとする各クラスタのタスクに御参加頂きたい.

謝辞

熱心に議論をしていただき，研究を強力に推進していただいた，NTCIR 質問応答関連タスクの参加者の皆様に感謝いたします.

また，テキストデータの研究利用をご承諾いただいた，毎日新聞社，読売新聞社をはじめとする関係各団体に感謝いたします.

参考文献

[1] NTCIR Project (NII Test Collection for IR Systems). http://research.nii.ac.jp/ntcir/.

[2] Text REtrieval Conference (TREC) Home Page. http://trec.nist.gov/.

[3] Welcome to Cross Language Evaluation Forum. http://www.clef-campaign.org/.

[4] Jun'ichi Fukumoto. Question Answering System for Non-factoid Type Questions and Automatic Evaluation based on BE Method. In *Proceedings of the Sixth NTCIR Workshop*, May 2007.

[5] Jun'ichi Fukumoto, Tsuneaki Kato, and Fumito Masui. Question Answering Challenge (QAC-1): An Evaluation of Question Answering Task at NTCIR Workshop 3. In *Proceedings of the Third NTCIR Workshop*, 2002.

[6] Jun'ichi Fukumoto, Tsuneaki Kato, and Fumito Masui. Question Answering Challenge for Five Ranked Answers and List Answers — Overview of NTCIR4 QAC2 Subtask 1 and 2 —. In *Proceedings of the Fourth NTCIR Workshop*, June 2004.

[7] Jun'ichi Fukumoto, Tsuneaki Kato, Fumito Masui, and Tatsunori Mori. An Overview of the 4th Question Answering Challenge (QAC-4) at NTCIR Workshop 6. In *Proceedings of the Sixth NTCIR Workshop*, May 2007.

[8] Tsuneaki Kato, Jun'ichi Fukumoto, and Fumito Masui. Question Answering Challenge for Information Access Dialogue — Overview of NTCIR4 QAC2 Subtask 3 —. In *Proceedings of the Fourth NTCIR Workshop*, June 2004.

[9] Tsuneaki Kato, Jun'ichi Fukumoto, and Fumito Masui. An Overview of NTCIR-5 QAC3. In *Proceedings of the Fifth NTCIR Workshop Meeting*, December 2005.

[10] NTCIR-5 CLQA Task Organizer. NTCIR-5 Evaluation Results: CLQA. http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings5/cdrom/CLQA1/, 2005.

[11] NTCIR-6 CLQA Task Organizer. NTCIR-6 CLQA Evaluation Results. http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings6/NTCIR/CLQA/clqa.html, 2007.

[12] NTCIR-6 QAC Task Organizer. NTCIR-6 Supplemental Material: Question Answering Challenge (QAC4). http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings6/NTCIR/NTCIR-6-QAC/index.html, 2007.

[13] NTCIR-7 ACLIA Task Organizers. NTCIR-7 ACLIA Wiki. http://aclia.lti.cs.cmu.edu/wiki/moin.cgi/Home.

[14] NTCIR3 QAC Task Organizer and Committee. NTCIR3 QAC Task Formal Run Package. http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings3/eval/qac/qac.html, 2002.

[15] NTCIR4 QAC Task Organizer. NTCIR4 QAC2 2nd Question Answering Challenge Task Evaluation Results. http://research.nii.ac.jp/ntcir-ws4/NTCIR4-EV-CD/QAC2/ntc4-qac2-eval.html, 2004.

[16] NTCIR5 QAC3 Task Organizer. NTCIR5 Question Answering Challenge 3 Data Set. http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings5/cdrom/QAC3/QAC3index.html, 2005.

[17] Yutaka Sasaki, Hsin-Hsi Chen, Kuang hua Chen, and Chuan-Jie Lin. Overview of the ntcir-5 cross-lingual question answering task (CLQA1). In *Proceedings of the Fifth NTCIR Workshop Meeting*, December 2005.

[18] Yutaka Sasaki, Chuan-Jie Lin, Kuang hua Chen, and Hsin-Hsi Chen. Overview of the NTCIR-6 Cross-Lingual Question Answering (CLQA) Task. In *Proceedings of the Sixth NTCIR Workshop*, May 2007.