A Question/Answer Typology with Surface Text Patterns

Eduard Hovy, Ulf Hermjakob, and Deepak Ravichandran

USC Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292-6695
Tel:+1-310-448-8711/8731
{hovy,ulf,ravichan}@isi.edu

ABSTRACT

In this paper we announce the release of ISI's QA Typology, which is being made available on the web to support the rapid construction of new QA systems. The Typology has been augmented with surface-level patterns associated with answer types, allowing systems to locate answers of the desired type in text by simple string matching. These patterns are extracted from the web automatically. We describe the process of their extraction, compression, and accuracy determination.

Keywords

Question answering; QA typology; answer patterns

1. QA TYPES AND TYPOLOGIES

The recent TREC-10 Question Answering competition included over 65 participating systems. The vast majority (if not all) performed some variant of answer typing, in which a process analyzes the input question to determine the desired type of the answer. Answer types are used by systems as a matching criterion to filter out candidate answers that look likely. For example, "Who is the richest person in the world?", "Where is Cambodia?", and "When did Marilyn Monroe marry Arthur Miller?" may have answer types *ProperPerson*, *Location*, and *Date* respectively.

The same answer type can be intended by various forms of question, as in "What is the name of the person who invented xeroxing?", "Who invented xeroxing?", and "Who was the inventor of xeroxing?". Similarly, the answer can occur in different forms, such as "Chester F. Carlson invented xeroxing", "the man who invented photocopying was Chester Carlson", and "Inventing the xeroxing process was a high point of Chester F. Carlson's life". What we called an answer type is thus a kind of equivalence class of all these phrasings, or a relation that links all the question forms to all their answer forms. We call this equivalence class a *Qtarget*; for this example it might be *Person*. Most QA systems include a list of such Qtargets, typically ranging in number from around 10 to around 50, and starting with *Who, When, Where,* and *What*.

Since many QA systems associate specific matching information (indicative words, surface word patterns, etc.) with their Qtargets, it is useful to create more specific alternatives that narrow the equivalent sets. Thus *Person* might be specialized to *Inventor*, and be associated with words such as "invent", "invention", "discover", "discovery", and "create". Other specializations of *Person* might be *Artist* (with "perform", "sing") and *Author* ("write", "book", "publish"). In addition, questions often explicitly require more or less specific answers ("Where is Vesuvius?" vs. "In which country is Vesuvius?")

The hierarchicalized Qtargets form a typology that defines the types of questions the system can handle. The hierarchicalization can be exploited for backoff matches, to allow more general Qtargets to apply in cases where specific ones fail. QA lists or typologies are reported in almost all QA system papers; see for example [4,1].

2. ISI'S QA TYPOLOGY

Over the past two years, we have created at ISI a QA Typology that currently contains 140 Qtargets. Our initial Typology of about 75 nodes was derived from an analysis by one of our students of over 17,000 questions, downloaded from answers.com [7,8]; see http://www.isi.edu/natural-language/projects/webclopedia/Taxonomy/taxonomy_toplevel.html.

Subsequently, we have been restructuring and extending the Typology to its current form.

Qtargets are of five types.

Abstract Qtargets. Some qtargets are not typical semantic types but are specific to QA. For example, the usual reading of "who was Mother Theresa?" is "why is the individual known as Mother Theresa famous?". The Qtarget *WhyFamous* is not a semantic class. These Qtargets are shown in Figure 1.

Semantic Qtargets. These Qtargets, the largest class, limit the search space to sentence constituents that satisfy a particular semantic class with respect to the Webclopdia ontology, which currently contains about 10,000 items, mostly extracted from WordNet. Some semantic Qtargets are shown in Figure 2.

Syntactic Qtargets. Other Qtargets apply when the system cannot determine a specific semantic type for the answer, but can specify the desired syntactic type. Syntactic qtargets are

```
Abstract qtargets
         WHY-FAMOUS
              WHY-FAMOUS-PERSON
              - Who was Jane Goodall?
              - What is Jane Goodall famous for?
         DEFINITION - What is platinum?
         ABBREVIATION-EXPANSION - What does NAFTA stand for?
         ABBREVIATION - What's the abbreviation for limited partnership?
         SYNONYM - Aspartame is also known as what?
         CONTRAST - What's the difference between DARPA and NSF?
         POPULATION - How many people live in Greater Tokyo?
         VERACITY
              YES-NO-QUESTION - Does light have weight?
              TRUE-FALSE-QUESTION - Chaucer was an actual person. True or false?
         PHILOSOPHICAL-QUESTION - What is the meaning of life?
```

Figure 1. QA typology: Abstract Qtargets.

```
Semantic qtargets
    TEMP-LOC - before the conference
         DATE - July 4
              DATE-WITH-YEAR - July 4, 1776
              ...
         YEAR-RANGE
              DECADE - in the 1980s
              ...
         TEMP-LOWITH-YEAR - July 1776
              DATE-WITH-YEAR
         DATE-RANGE - July 12-14
         TIME - at 2:15pm EST
    AT-LOCATION - under the bed
    PROPER-NAMED-ENTITY
         PROPER-PERSON - John F. Kennedy
         PROPER-LANGUAGE - Afrikaans
         PROPER-ANIMAL - Lassie
         PROPER-PLACE
              CONTINENT - Africa
              WORLD-REGION - Middle East
              ...
              PROPER-COUNTY - Orange County
              PROPER-CITY - Chicago
              PROPER-BODY-OF-WATER
                   PROPER-OCEAN
                   ...
              PROPER-ISLAND
              PROPER-MOUNTAIN
         ...
              PROPER-STAR-CONSTELLATION - Orion
              ...
              PROPER-AMUSEMENT-PARK - Disneyland
              PROPER-HOTEL - Holiday Inn
              PROPER-UNIVERSITY - Harvard
              ...
              PROPER-AIRPORT - LAX
         PROPER-ORGANIZATION
              PROPER-SPORTS-TEAM
                   ...
              PROPER-POLITICAL-PARTY
```

```
          PROPER-COMPANY
                PROPER-BROADCASTING-COMPANY
                ...
          GOVERNMENT-AGENCY - House of Representatives, OSHA
PLANT-FLORA
     FLOWER
     TREE
SUBSTANCE
     SOLID-SUBSTANCE
          METAL
     LIQUID
          BEVERAGE
QUANTITY
     MONETARY-QUANTITY - How much does one ton of cement cost?
     SPATIAL-QUANTITY
          DISTANCE-QUANTITY
          - How far is it from Denver to Aspen?
          - How tall is the Sears Building?
          AREA-QUANTITY - How big is Australia?
          VOLUME-QUANTITY - How much milk did you buy?
     TEMPORAL-QUANTITY
     - How long did Desert Storm last?
     - What is the life expectancy for crickets?
     SPEED-QUANTITY - How fast is sound?
     ACCELERATION-QUANTITY - What's the gravity on the moon?
     NUMERICAL-QUANTITY/I-ENUM-CARDINAL
          - How many moons does Jupiter have?
     FREQUENCY-QUANTITY
     - How often does Old Faithful erupt at Yellowstone National Park?
     SCORE-QUANTITY
     - What was the score of the World Cup final Germany vs. Netherlands?
     PERCENTAGE - What is the sales tax rate in New York?
     TEMPERATURE-QUANTITY - How hot is the core of the earth?
     MASS-QUANTITY - What is the average weight of a Yellow Labrador?
     POWER-QUANTITY
     - How many megawatts will the new power plant in Indonesia produce?
     ...
C-LOCATOR
     C-PHONE-NUMBER, C-ADDRESS, C-ZIP-CODE, C-EMAIL-ADDRESS, C-URL
UNIVERSITY-AGENCY - biology department
SPIRITUAL-BEING
OCCUPATION-PERSON - computer scientist
ANIMAL
HUMAN-FOOD
BODY-PART
TEMPORAL-INTERVAL
     DAY-OF-THE-WEEK, MONTH-OF-THE-YEAR, SEASON ...
DISEASE - arthritis
INSTRUMENT - hammer
MUSICAL-INSTRUMENT -trumpet
SPORT - football
LEFT-OR-RIGHT - left, right
I-EADJ-COLOR - blue
I-EADJ-NATIONALITY - German
```

Figure 2. QA Typology: Some Semantic Qtargets.

fairly weak, in that they generally don't restrict the search space much. Webclopedia uses *S-NP* as the default Qtarget. The four syntactic Qtargets are:

```
S-NP, S-NOUN
   What does Peugeot manufacture?
S-VP
   What did John Hinckley do to impress Jodie
   Foster?
S-PROPER-NAME
```

Role Qtargets. These Qtargets specify constituents of the parse tree of the question and candidate answer:

```
ROLE REASON
  Why did David Koresh ask the FBI for a word
  processor?
ROLE MANNER
  How did David Koresh die?
```

For example, in the (simplified) parse tree

```
The tournament was cancelled due to bad
   weather.
 ((SUBJ LOG-OBJ) [2] The tournament
  (PRED) [5] was cancelled
  (REASON) [6] due to bad weather
  (DUMMY) [14] .
 )
```

the phrase "due to the bad weather" satisfies the Qtarget *Role Reason*. This constraint is independent of the syntactic category, which also could have been a subordinate clause ("because the weather was so bad") or a verb phrase ("to avoid injuries").

Slot Qtargets. Slot Qtargets refer to non-syntactic information associated with sentence constituents. Slots may be filled during parsing or later. Some examples are:

```
SLOT TITLE-P TRUE
   Name a novel written by Proust.
SLOT QUOTE-P TRUE
   What did Richard Feynman say upon hearing he
   would receive the Nobel Prize in Physics?
SLOT POSSIBLE-REASON-P TRUE
```

ISI's QA systems Webclopedia [8] and Textmap [9] both employ the Typology. Both systems can combine Qtargets, using variable strengths:

Question: Where is the Getty Museum?

Qtarget: ((*Proper-city* 1.0) (*At-location* 0.7)
(*Proper-place* 0.7) …)

indicating that the system would prefer a proper city, but could accept something tagged by the named entity tagger just as a general location, or as a place with a name [5].

3. ANSWER PATTERNS

At the recent TREC conference, several systems emphasized the value of a surface-oriented pattern matching approach to QA. The Insight system from Moscow [11] used some hundreds of surface-level patterns to identify answer strings without (apparently) applying Qtargets or similar reasoning. For example, for *BirthYear* questions such as "which year was Mozart born?" the phrase "Mozart (1756 – 1791)…" provides the answer using the general template

NAME_OF_PERSON (BIRTHYEAR – DEATHYEAR)

Several other systems also defined word-level patterns indicating specific Qtargets; e.g., [10]. The Microsoft system [3] extended the idea of a pattern to its limit, by reformulating the input question as a declarative sentence and then retrieving the sentence verbatim, with its answer as a completion, from the web using the normal search engines. For example, "who was Chester F. Carlson?" would be transformed to "Chester F. Carlson was" and submitted. Although this approach might yield many wrong answers (including "Chester F. Carlson was born February 8, 1906, in Seattle"), the sheer number of correct answers often wins the day.

We estimate that word-level patterns can provide at least 25% of the MRR score defined for TREC (although some systems claimed considerably higher results; see [11] and discussion in [6]). In order to determine their power and reap their benefits, we collected all the patterns associated with as many Qtargets as made sense (some Qtargets, such as *Planets* and *Oceans*, are known closed sets that require no patterns).

We developed an automated procedure to learn such patterns from the web, using Altavista (because it returns 1000 documents per query), and to measure their Precision. More formally this experiment can be phrased as "Given a QA pair such as (NAME_OF_PERSON BIRTHYEAR), extract from the web all the different patterns (TEMPLATEs) that contain this QA pair along with the precision of each pattern". We have inserted into the Typology the patterns for approx. 20 Qtargets, recording their Precision scores and relative frequencies of appearance.

The procedure contains two parts:

1. Extracting the patterns
2. Calculating the precision of each pattern

Algorithm 1: Extracting patterns

1. An example of the question-answer pair for which the pattern is to be extracted is passed to a search engine. To learn the pattern for the pair (NAME_OF_PERSON BIRTHYEAR) we submit the query "Gandhi 1869" to Altavista.

2. The top 1000 documents returned by the search engine are retrieved.

3. These documents are broken into sentences by a simple sentence breaker. Only sentences that contain both the Question and the Answer term are retained. BBN's IdentiFinder named entity tagger [2] is used to remove the variations caused by writing a name or a date in different forms.

5. Each of these sentences is converted into a Suffix tree, to collect counts on all phrases and subphrases present in the document.

6. The phrases obtained from the Suffix tree process are filtered so that only those containing both the Question and the Answer terms are retained. This yields the set of patterns for the given QA pair.

Algorithm 2: Calculating the precision of each pattern

1. The Question term alone (without the Answer term) is given as query to Altavista.

2. As before, the top 1000 documents returned by the search engine for this query are retrieved.

3. Again, the documents are broken into sentences.

4. Only those sentences with the Question term are saved.

5. For each pattern obtained in step 6 of Algorithm 1, a pattern-matching check is done against each sentence obtained from step 4 here, and only the sentences containing the Answer are retained. This is used to calculate the precision of each pattern according to the formula

   Precision = # patterns matching the Answer
   / total # matches for every pattern

6. Only those patterns are retained for which sufficient examples are obtained in step 5.

To increase the size of the data, we apply the algorithms with several different examples of the same Qtarget. Thus in Algorithm 1 for *BirthYear* we used Mozart, Gauss, Gandhi, Nelson Mandela, Michelangelo, Christopher Columbus, and Sean Connery, each with its birth year. We then applied Algorithm 2 with just these names, counting the yields of the patterns on the exact birth years (no additional words or reformulations, which would increase the yield score).

The results were quite good in some cases. For the rather straightforward *BirthYear* patterns are:

Precision	**#Correct**	**# Found**	**Pattern**
1.0	122	122	<NAME> (<BD> - <DD>
1.0	15	15	<NAME> (<BD> - <DD>) ,
1.0	13	13	, <NAME> (<BD> - <DD>)
0.9166	11	12	<NAME> was born on <BD> in
0.9090	10	11	<NAME> : <BD> - <TIME>
0.6944	25	36	<NAME> was born on <BD>

Note the overlaps among patterns. By not compressing them further we can record different precision levels.

The *Definition* Qtarget posed greater problems. E.g., disease (the names *jaundice, measles, malaria, and tuberculosis* were paired with the term *disease*, but not also with *illness, ailment,* etc., although this would have increased the yield):

1.0	46	46	heart <TERM> , <NAME>
1.0	35	35	<NAME> & tropical <TERM> weekly
1.0	30	30	venereal <TERM> , <NAME>
1.0	24	24	lyme <TERM> , <NAME>
1.0	22	22	, heart <TERM> , <NAME>
1.0	21	21	' s <TERM> , <NAME>
0.9565	22	23	lyme <TERM> <NAME>
0.9	9	10	s <TERM> , <NAME> and
0.8815	67	76	<NAME> , a <TERM>
0.8666	13	15	<TERM> , especially <NAME>

4. EXTENDING THE TYPES OF TYPES

This work has necessitated extensions to the original Typology. Because the Typology initially focused on matching of answers, most Qtargets express requirements on a single entity: the answer. In this regard the Typology resembles the typical question type hierarchies of QA systems.

The patterns, however, represent relationships between the two anchor points of each pattern. As a result, patterns pick out subsets of Qtargets and cannot easily be incorporated into the Typology. For example, the general Qtarget *Date* has to be split up into *BirthDate*, *DeathDate*, *WeddingDate*, *DiscoverDate*, *GraduationDate*, *ConstructionDate*, and so on. All the generalizations captured in the Typology under *Date* about the different forms in which dates can be expressed (day+month+year, month+year, year, etc.) have to be repeated for each pattern Qtarget. Naturally, this forces the patterns to contain not only surface forms (words and punctuation, but also type markers (*Date, NumericalAmount, MoneyAmount...*). Work is underway to automatically recognize occurrences of such expressions and replace them with type markers.

5. CONCLUSIONS

Overall, we are satisfied with the coverage of the Typology, and with its utility in Webclopedia's QA processing. A survey of the papers in the TREC 9 and 10 proceedings shows that most systems tend to use more than a dozen Qtargets, while only one system [4] uses in the order of 1000 (in essentially a semantic answer type ontology).

By making public the QA Typology and its patterns we hope to enable a wider range of people to build basic QA systems quickly, and then to help investigate the question of QA.

6. REFERENCES

[1] Abney, S., M. Collins, and A. Singhal. 2000. Answer Extraction. *Proceedings of the Applied Natural Languge Processing Conference (ANLP)*. Seattle, WA (296–301).

[2] Bikel, D., R. Schwartz, and R. Weischedel. 1999. An Algorithm that Learns What's in a Name. *Machine Learning—Special Issue on NL Learning*, 34, 1–3.

[3] Brill, E., J. Lin, M. Banko, S. Dumais, and A. Ng. 2001. Data-Intensive Question Answering. *Proceedings of the TREC-10 Conference*. NIST, Gaithersburg, MD (183–189).

[4] Harabagiu, S., M. Pasca, and S. Maiorano. 2000. Experiments with Open Domain Textual Question Answering. *Proceedings of the 18th COLING Conference*. Saarbrücken, Germany (292–298).

[5] Hermjakob, U. 2001. Parsing and Question Classification for Question Answering. In *Proceedings of the Workshop on Question Answering at the Conference ACL-2001*. Toulouse, France.

[6] Hermjakob, U. 2002. In prep.

[7] Hovy, E.H., L. Gerber, U. Hermjakob, M. Junk, and C.-Y. Lin. 2000. Question Answering in Webclopedia. *Proceedings of the TREC-9 Conference*. NIST. Gaithersburg, MD.

[8] Hovy, E.H., U. Hermjakob, and C.-Y. Lin. 2001. The Use of External Knowledge in Factoid QA. *Proceedings of the TREC-10 Conference*. NIST, Gaithersburg, MD (166–174).

[10] Oh, JH., KS. Lee, DS. Chang, CW. Seo, and KS. Choi. 2001. TREC-10 Experiments at KAIST: Batch Filtering and Question Answering. *Proceedings of the TREC-10 Conference*. NIST, Gaithersburg, MD (354–361).

[11] Soubbotin, M.M. and S.M. Soubbotin. 2001. Patterns of Potential Answer Expressions as Clues to the Right Answer. *Proceedings of the TREC-10 Conference*. NIST, Gaithersburg, MD. (175–182)