D4: QA SYSTEM

Maria Alexandropoulou

Max Kaufmann

Alena Hrynkevich

QA System

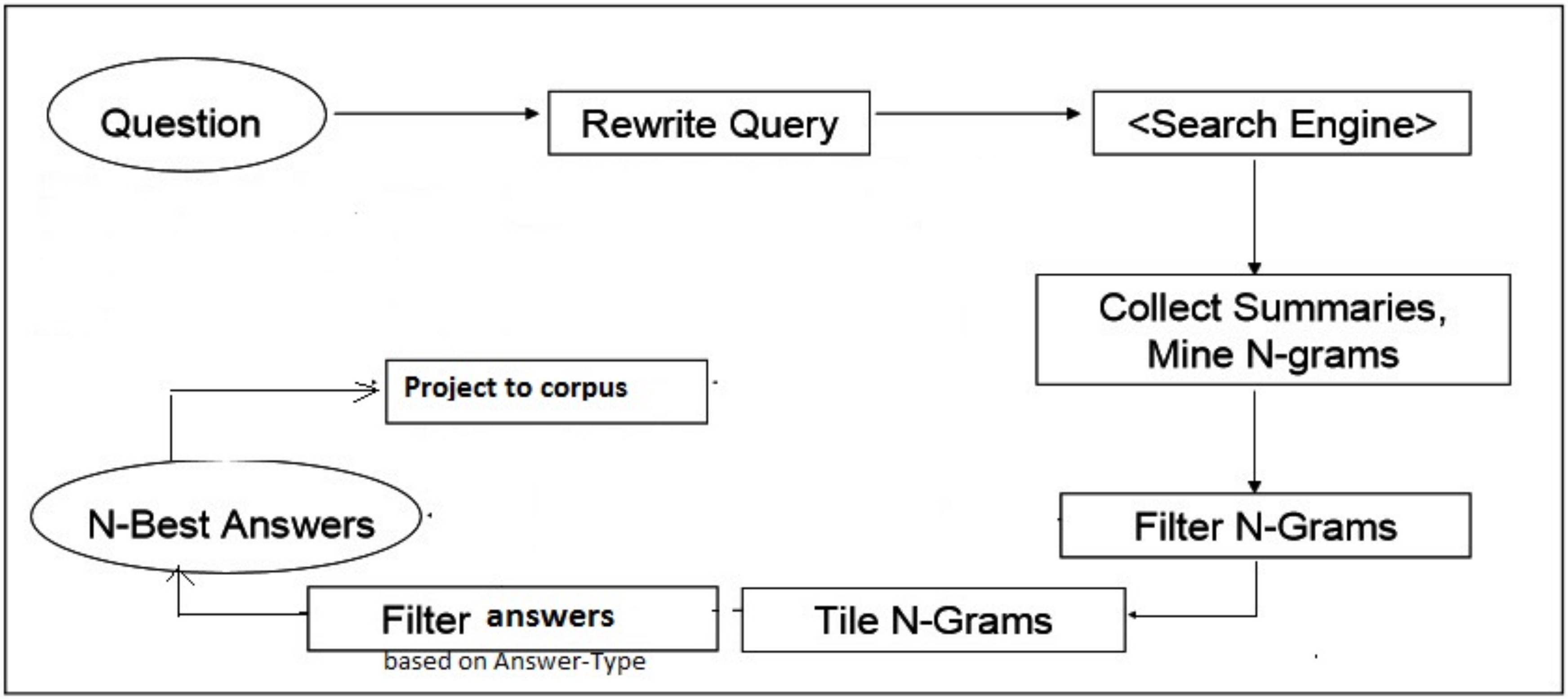

Fig. 1 ARANEA-inspired QA-System architecture

D4 Answer Extraction Improvements

Web summaries collection

- Run Stanford NER for specific question types
- Extract NE of a given type instead of whole summary

(ex. for HUM:ind -> PERSON_ or NUM:date -> DATE_)

Question Type	Stanford NER
HUM:ind	PERSON_
LOC:city LOC:country LOC:state LOC:other	LOCATION_
NUM:money	MONEY_
HUM:gr ABBR:abb ABBR:exp	ORGANIZATION_
NUM:money	MONEY_
NUM:perc	PERCENT_
NUM:date	DATE_

Fig. 2 Mapping between q classes and NE

D4 Answer Extraction
N-gram processing

- Collection of N-grams (**N = 1,2,3,4,5**)
- Initial weighting depending on query type (**1:3**)
- Scoring of compound n-grams
- Heuristic-based filtering
- Re-scoring
- Sorting
- **Tiling**
- **Class-based filtering**

Tiling

- merges similar answers
- assembles longer answers from overlapping smaller answer fragments

BEFORE	AFTER
possibility of Tebow	rule out the possibility of Tebow becoming a successful
possibility of Tebow becoming	believes that Tim Tebow Hall of Fame QB In CFL
Tebow Hall	suggested that perhaps Tim Tebow pursue a career in another
Tim Tebow	modern era to be inducted into the Pro Football Hall of Fame in 2006
Tebow pursue a career	told radio station KILT-AM in If you can't throw
Tebow	Canadian Football Hall of Fame quarterback in the modern era
Tebow becoming a successful	Hall of Famer
Hall of Fame QB	Fame and the Pro Football Talk
Tebow becoming	ESPN Pro Football League
believes that Tim Tebow	inducted into the pro
Tebow pursue	NFL Pro Bowl
Hall of Fame quarterback	became the first Black quarterback in the modern

Final Class-Based Filtering

- Gazetteers: countries, states, cities
- Person names (rule-based)
- Dates normalization (based on answer, question and question topic)

- References
- E. Brill, J. Lin, M. Banko, S. Dumais and A. Ng, Data-intensive question answering, 2001
- J. Lin, An Exploration of the Principles Underlying Redundancy-Based Factoid Question Answering, 2007
- X. Li, D. Roth, Learning Question Classifiers, 2002

Answer Projection

- Four ways were used for answer projection (Mishne, G. & De Rijke (2005)) :
  - Question and answer terms are required within a window of 15 words.
  - Question and answer terms are required as above without the limitation of being in a span of 15 words.
  - Question terms and answer as a phrase are searched for (not limited span)
  - Question terms and boosted answer terms with the answer terms required within a certain window.

Answer Projection

Approach	Strict Score
Question/Answer terms within a span	0.0737
Question/Answer terms	0.0860
Question terms and answer as a phrase	0.0821
Question terms and boosted answer terms. Answer within a span.	0.0802

System Results

system	lenient	strict
Baseline	0.0493692039364	0.0109649916451
Dev set	0.289616571992	0.0883386832141
Test set	0.213500927385	0.0528715106732

Fig. 4 Evaluation results

Issues and Successes

- Great score improvement compared to baseline!
- **BUT:**
- maybe not good enough coverage in question reformulations for some types of questions;
- definitely poor Bing results even for good queries;
- some issues with classification of questions: misclassifications, lack of support of some specific classes from eval set ex. City+State
- missing rules in filters

- We still need to investigate the reasons for pretty big gaps between lenient and strict scores

Semantic Role Labeling

- Inspired by “Using Semantic Roles to Improve Question Answering”
- Used off-the-shelf tool: SEMAFOR
  - Unable to incorporate into pipeline (yet)
  - Too slow for realtime
- Qualitative analysis of SEMAFOR showed good SRL results
- Studied the interaction between question class and semantic roles (FrameNet)
- Question classifier from D3 got ~90% accuracy on test data, so for this experiment we assume that classifier-generated labels are accurate

Top frames for question classes

- NUM:date (63)
  - Temporal_collocation: 32
  - Calendric_unit: 31
  - Intentionally_act: 26
- NUM:count (98)
  - Quantity: 97
  - Intentionally_act: 39
  - Cardinal_numbers: 31
- LOC:country (13)
  - Political_locales: 13
  - Natural_features: 2
  - Intentionally_act: 2

Top frames for question classes

- ENTY:color (1)
  - Wearing: 1
  - Dressing: 1
  - Race_descriptor: 1
- OTHER:class (174)
  - Age: 47
  - Temporal_collocation: 39
  - Leadership: 37

Analysis

- Question classes about locations line up extremely well with the Political_locales frame
- Most question classes lined up extremely well with 1-3 Frames
- Few examples with strong alignments, but not enough data

Frame Elements and Question class

- Frame elements are useful because they might be able to tell us what a question is asking for, not just what it's about
- Soft approach in Shen & Lapata (2007) is ideal here, but requires lots of work to implement

Top frame elements for question classes

- NUM:date (63)
  - Act/Intentionally_act: 26
  - Relative_time/Calendric_unit: 26
  - Child/Being_born: 10
- LOC:country (13)
  - Locale/Political_locales: 13
  - Locale/Natural_features: 2
  - Name/Natural_features: 2
- NUM:count (98)
  - Quantity/Quantity: 97
  - Individuals/Quantity: 85
  - Act/Intentionally_act: 39

Analysis

- Frame Element analysis is not as generalizable, because 1 question class can ask for many different types of things
  - HUM:gr (10)
    - Business/Businesses: 5
    - Owner/Possession: 2
    - Intoxicant/Intoxicants: 2
- As Shen and Lapata (2007) pointed out, there are coverage gaps in FrameNet

Future Work

- Question classification
  - Use SRL features for better question classification
  - Try with all 50 question classes, instead of reduced set
- Ranking
  - If we see that a frame frequently appears in a question class, rank answers higher if they have that frame
    - Easy example: NUM:money class and Money frame.
- Answer phrase extraction
  - Use semantic structure matching to extract expected answer phrase.