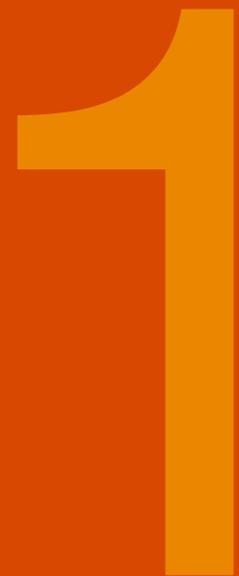# A Recall Oriented Approach to Question Answering

Valentin Jijkoun, Gilad Mishne, Maarten de Rijke,
Stefan Schlobach, David Ahn, Karin Müller

Informatics Institute
University of Amsterdam

1

Some Observations

- ▶ The vocabulary gap
  - ◾ identifying answer documents
  - ◾ question and answers phrased in different terms
  - ◾ a "recall" problem
- ▶ Different strategies bridge gap in different ways
  - ◾ pattern-based
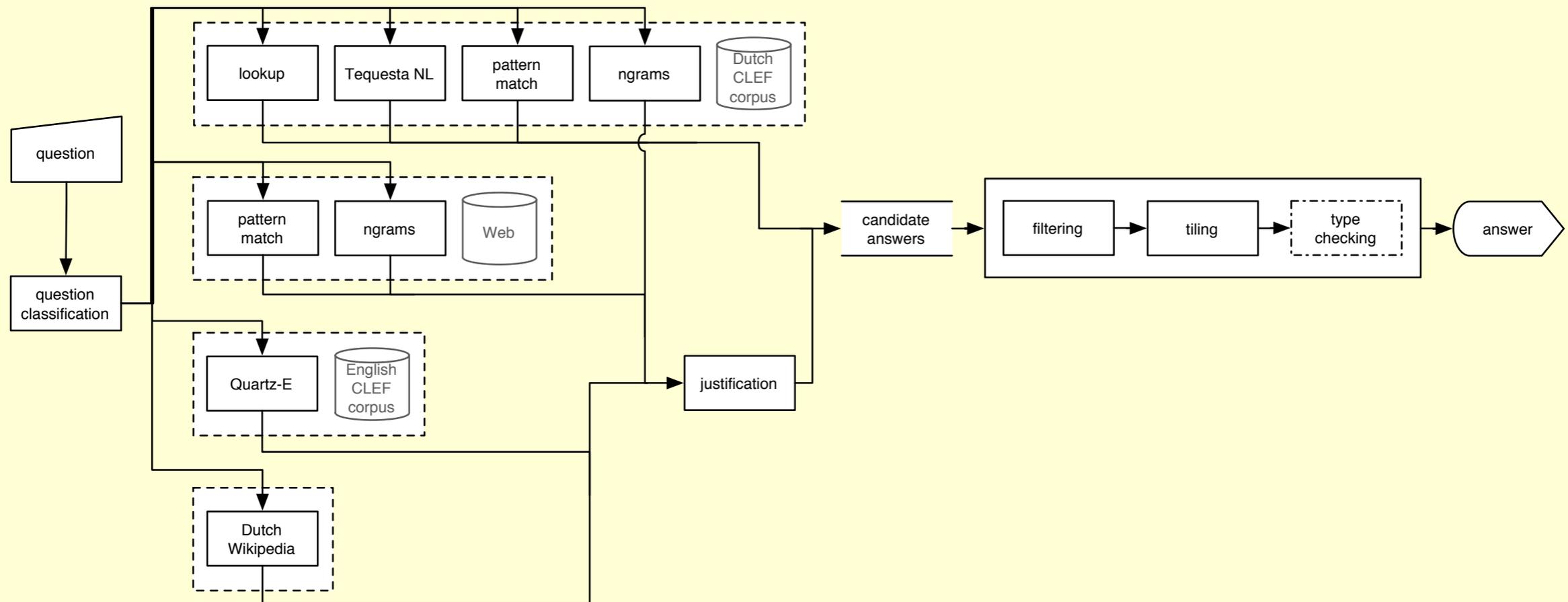 - ◾ redundancy-based
  - ◾ knowledge-intensive

Some Observations

wers

run	Definition (#)		Factoid (#)								Total	
	org [11]	per [9]	loc [28]	man [15]	mea [20]	obj [12]	org [20]	oth [27]	per [28]	tim [30]	# [200]	%
bgas041bgen	2	3	5	2	1	2	1	2	4	4	26	13.00
dfki041deen	4	0	10	2	2	1	5	5	6	12	47	23.50
dltg041fren	3	3	8	5	2	1	1	2	4	9	38	19.00
dltg042fren	3	3	4	3	1	1	1	2	3	8	29	14.50
edin041deen	1	3	6	2	0	0	2	2	4	8	28	14.00
edin041fren	0	1	7	3	1	1	2	4	3	11	33	16.50
edin042deen	1	4	6	4	1	2	2	5	3	6	34	17.00
edin042fren	0	3	7	4	3	1	2	4	4	12	40	20.00
hels041fien[11]	0	0	3	0	2	0	5	4	5	2	21	10.88
irst041iten	0	5	11	0	1	0	6	3	8	11	45	22.50
irst042iten	0	5	5	0	1	0	2	5	6	11	35	17.50
lire041fren	3	1	9	0	1	0	3	0	1	4	22	11.00
lire042fren	2	1	13	0	1	0	4	1	6	11	39	19.50
combination	7	5	26	6	7	5	18	10	22	24	130	65.00

Table 5: Results of the bilingual English runs, according to answer types of questions.

g to answer types of questions.

A Multi-Stream Approach to Question Answering

- Run many multiple strategies for identifying candidate answers

- Supplement with filtering mechanism to weed out "obviously" incorrect answers

- Have a voting mechanism, conditioned on question type, stream, source, … to select "the answer"/ rank the answers

Quartz-N-2004
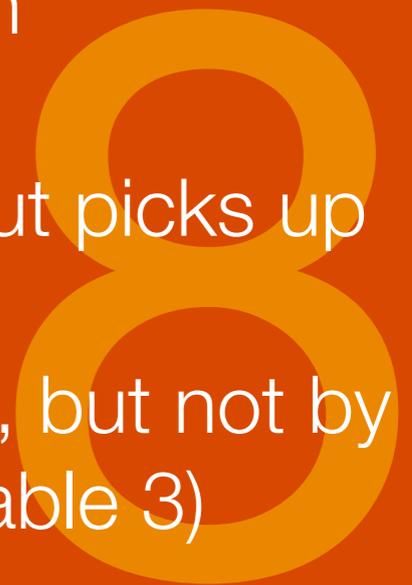
Changes since CLEF 2003

- ▶ Follow-up to detailed error analysis post-CLEF 2003

- ▶ More streams
  - ◾ table stream
  - ◾ Dutch Wikipedia

- ▶ Better modules
  - ◾ question classification
  - ◾ retrieval
  - ◾ justification
  - ◾ voting
  - ◾ filtering, type checking

6

Runs

- ▶ Submitted three runs
  - ■ monolingual NL→NL
  - ■ monolingual NL→NL with additional filtering and sanity checking
  - ■ bilingual EN→NL, using Systran for translating questions
- ▶ How much does the filtering help?
- ▶ How much does the translation hurt?

Run	Accuracy	Acc. F	Acc. D	NIL Prec	NIL Rec
uams041nlnl	44.0	42.4	56.5	0.0	0.0
uams042nlnl	45.5	45.2	47.8	0.56	0.25
uams041ennl	35.0	31.1	65.2	0.0	0.0

▶ Filtering seems to help factoids
▶ But definitions seem to benefit from lenient approach
▶ Filtering helps NIL scoring
▶ Bilingual run performs less than monolingual runs, but picks up answers not found by the monolingual runs
▶ 25 questions answered correctly by the bilingual run, but not by either of the monolingual ones… (Overview paper, Table 3)

8

Examples

- Disappointments
  - Wat is de hoofdstad van Rusland?
    - Minsk
  - Izmir ligt in welk land?
    - Ankara
  - Met wie is Whoopi Goldberg getrouwd?
    - always on the move
  - Bij welke basketballploeg speelt Shaquille O'Neal?
    - vijf jaar

Examples

- ► Entertaining

  - ■ What color hair did Thomas Jefferson have before grey?

    - ► green

  - ■ How many engines does a plane have?

    - ► 2002

  - ■ Wat is het symbool van de leider van de Giro d'Italia?

    - ► Forza Italia

10

Examples

- ▶ When things go right
  - ▪ Hoe stierf Jimi Hendrix?
    - ▶ overdosis slaaptabletten
  - ▪ Wat is FARC?
    - ▶ Colombiaanse linkse guerrilla-organisatie
  - ▪ Wie is Marvin Minsky?
    - ▶ geleerde op het gebied van de kunstmatige intelligentie

Conclusions

- ▶ Lessons learned
  - once again, multi-stream approach proved reasonably effective
  - doing a thorough error analysis pays off
  - RTFO
- ▶ Next steps
  - move more analysis off-line
  - additional filtering/type checking
  - exploit external resources

12

Join the "new veterans":
Online demo available at http://ilps.science.uva.nl
30 month postdoc position in IR available