Exploring the Effects of Root Expansion, Sentence Splitting and Ontology on Arabic Answer Selection

Ahmed Magdy Ezzeldin,   Yasser El-Sonbaty,   Mohamed Hamed Kholief
College of Computing and Information Technology, AASTMT
**Alexandria, Egypt**

Outline

- Arabic QA and its significance
- QA pipeline
- Importance of answer selection to QA
- QA4MRE (Question Answering for Machine Reading Evaluation)
- Motivation
- Objectives
- Test-set & Evaluation metrics
- Related Works
- Tools and resources
- Proposed Approach
- Evaluation and discussion
- Future work

Arabic QA and its significance

- **Question Answering** (QA) : automatically providing an answer for a question posed by a human in a natural language

- **Significance** of Arabic QA

  – **Arabic** has more than 350 million native speakers

  – A lot of Arabic **content** on the Internet

  – High need for fast precise **information**

  – IR based approaches do **not satisfy** this need

QA pipeline
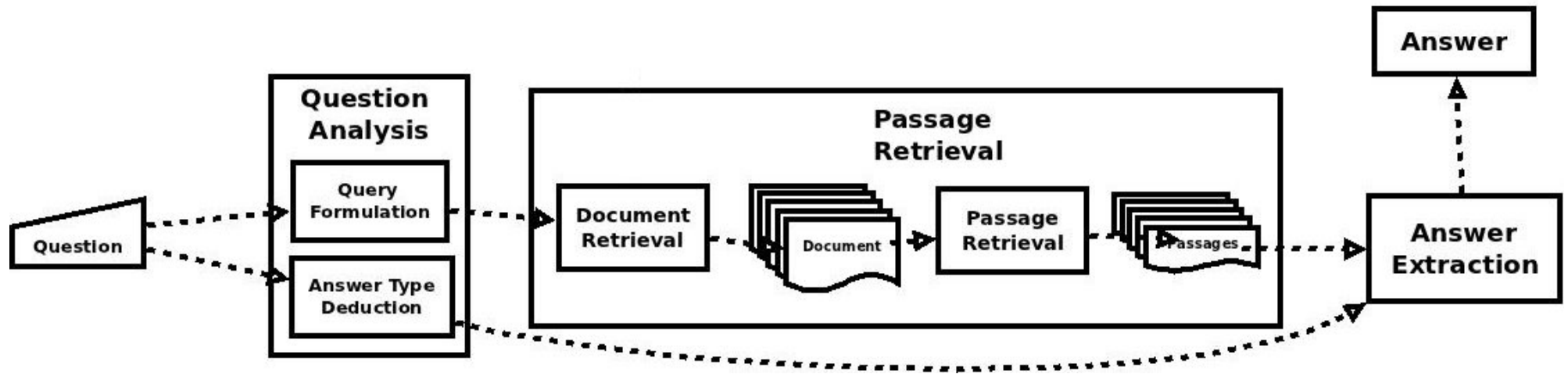
- Question Analysis
- Passage Retrieval
- Answer Extraction

Importance of Answer Selection to QA

- **Error propagation** in QA pipeline makes systems hit an upper bound of 60%

- Most QA systems provide **5 answer choices** for each question and most of the time the answer is not in the first choice

- Systems should be **certain** about the questions they answer

QA4MRE

- Question Answering for Machine Reading Evaluation

- Skips the answer generation tasks of QA (**Passage Retrieval**)

- Focuses only on the **answer selection** and validation subtasks

- A typical QA4MRE system chooses 1 answer form **5 choices** supported with 1 document that contains the answer

- QA4MRE is used **interchangeably** with Answer Selection and Validation

Motivation

- Create an efficient answer selection and validation module to:
  - Improve the **performance** of any Arabic QA system
  - **Reduce** the effect of **error** that propagates through the QA pipeline by selecting the correct answers only
  - Enhance the **certainty** of Arabic QA systems

Objectives

- Mimic the human behavior by **analyzing** the reading **test document** instead of just analyzing the question

- Apply **syntactic & semantic** analysis and expansion to the reading test document and the question

- Make use of **background knowledge** to gain more contextual knowledge about the test document and question keywords

Test-set

- The QA4MRE test-set is composed of:
  - 4 topics ("AIDS", "Climate change", "Music and Society", and "Alzheimer")
  - 16 test documents (4 documents per topic)
  - 160 questions (10 questions per document)
  - 800 answer choices/options (5 per question)
  - 4 background collections of documents (1 per topic)
- Various question types
  - Factoid: (where, when, by-whom)
  - Causal: (what was the cause/result of event X?)
  - Method: (how did X do Y? or in what way did X come about?)
  - Purpose: (why was X brought about? or what was the reason for doing X?)
  - Which is true: (what can a 14 year old girl do?)

Evaluation Metrics

- Accuracy & **C@1** (Correctness at 1)
- C@1 is introduced in **CLEF** (Conference and Labs of the Evaluation Forum) 2011 by Penas et al.
- Gives **partial credit** for systems that leave some questions unanswered in cases of uncertainty

$$Accuracy = \frac{tp+tn}{tp+fp+tn+fn}$$

Where:

tp: True Positives

tn: True Negatives

fp: False Positives

fn: False Negatives

$$C@1 = \frac{1}{n}\left(n_R + n_U \frac{n_R}{n}\right)$$

Where:

n_R: number of correctly answered questions

n_U: number of unanswered questions

n: total number of questions

Related Works

- The 2 Arabic QA4MRE systems in CLEF 2012
    - Trigui et al. 2012
    - Abouenour et al. 2012 (IDRAAQ)

- Best performing QA4MRE system on the same test-set uses the English test-set
    - Bhaskar et al. 2012 (**English** QA4MRE)

- Our first attempt in Arabic QA4MRE: ALQASIM 1.0 (Ezzeldin et al. 2013)

Related Works [continued]

- Abouenour et al. 2012 (IDRAAQ)

  – Accuracy : **0.13**

  – C@1 : **0.21**

- Used **JIRS** (Java Information Retrieval System) for passage retrieval (Distance Density N-gram Model)

- Semantic expansion using Arabic WordNet (**AWN**)

- Did **not** use the CLEF **background** collections

Related Works [continued]

- Trigui et al. 2012:
  - Accuracy : **0.19**
  - C@1 : **0.19**
- Has **not** marked any questions as **unanswered**
- **Retrieves** the passages that have the question keywords and **aligns** them with the answer choices
- **Finds** the best answer choice in the retrieved passages
- Expands the answer choices that could not be found in the retrieved passages using some inference rules on the **background collection**
- Depends on the background collection as it offers enough **redundancy** for the passage retrieval module

Related Works [continued]

- Bhaskar et al. 2012 (on the **English** test-set)
  - Accuracy : **0.53**
  - C@1 : **0.65**
- Combines each answer choice with the question in a **hypothesis**
- **Searches** for the hypothesis keywords in the document
- Uses **textual entailment** to rank the retrieved passages

Related Works [continued]

- **ALQASIM 1.0 (Ezzeldin et al. 2013)**
  - **Document Analysis**
    - Used MADA+TOKAN for **PoS** tagging and **stemming**
    - Semantic Expansion using AWN (**synonyms** only)
    - **Keyword** weights according to their **repetition** (more repetition is less likely to mark a question or answer snippet)
  - **Locating Questions & Answers**
    - **Keywords distances**
    - Return high **scoring snippets** for each Q and A
  - **Answer Selection**
    - According to **distance** between **Q and A** and their scores

Related Works [continued]

- The 2 Arabic QA4MRE systems

  - Use the **typical QA** pipeline that depend mainly on passage retrieval

  - Do **not analyze** the reading test **document**

- The **English** system (Bhaskar et al. 2012)

  - Analyzes the reading test **document**

  - Uses many English specific NLP **tools** like: semantic parsing, semantic role labeling, dependency analysis

- ALQASIM 1.0 (Ezzeldin et al. 2013) analyzes the document but does not use sentence splitting, root expansion or ontology based expansion.

Tools and resources

- **MADA+TOKAN** (Morphological Analyzer)
  - PoS tags, gender, number & light stem
- Arabic WordNet (**AWN**)
  - Semantic expansion by (synonyms)
- **Root** Stemmers
  - Khoja
  - **ISRI**
  - Tashaphyne

High Level Architecture
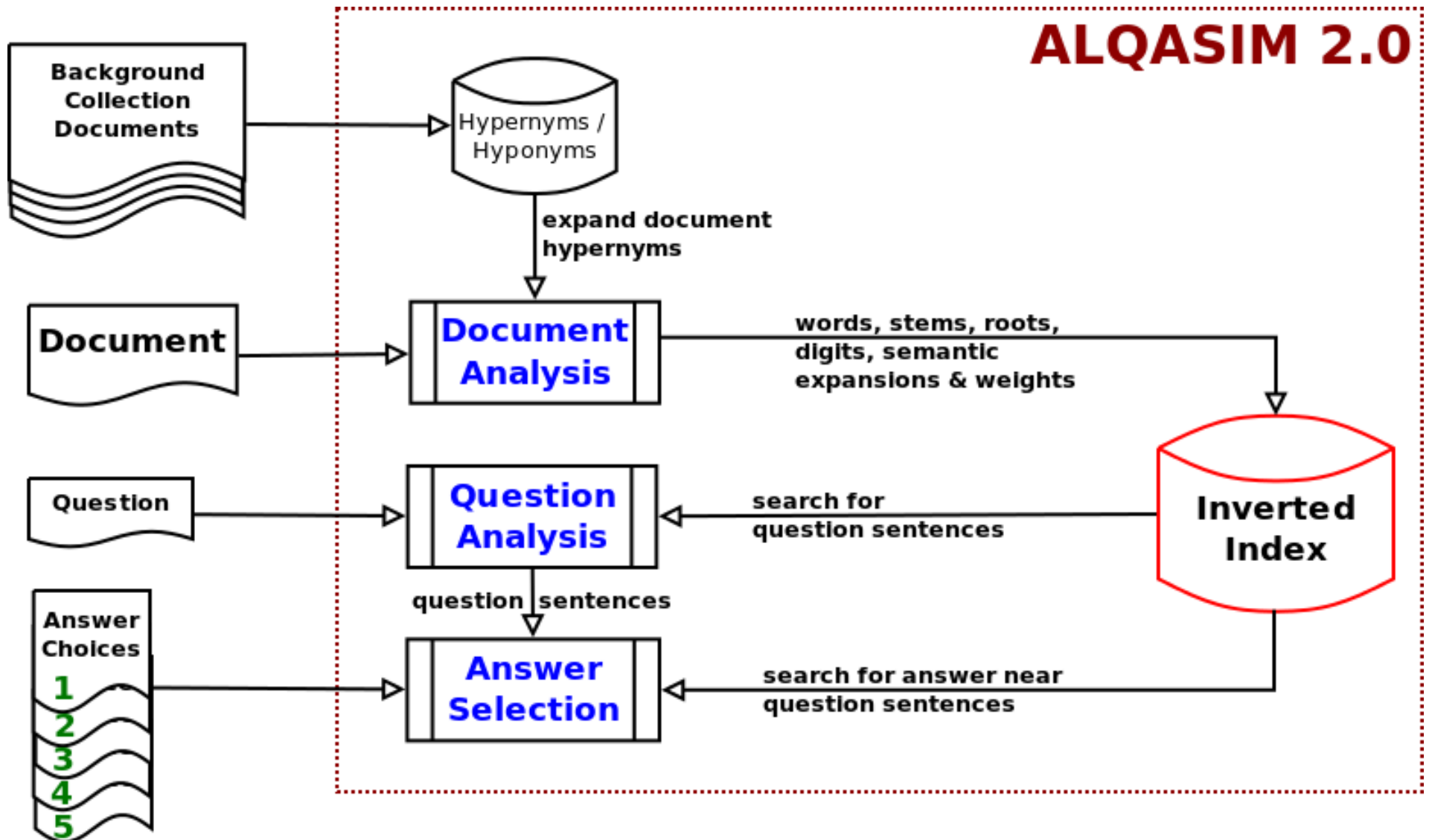
Document Analysis Module
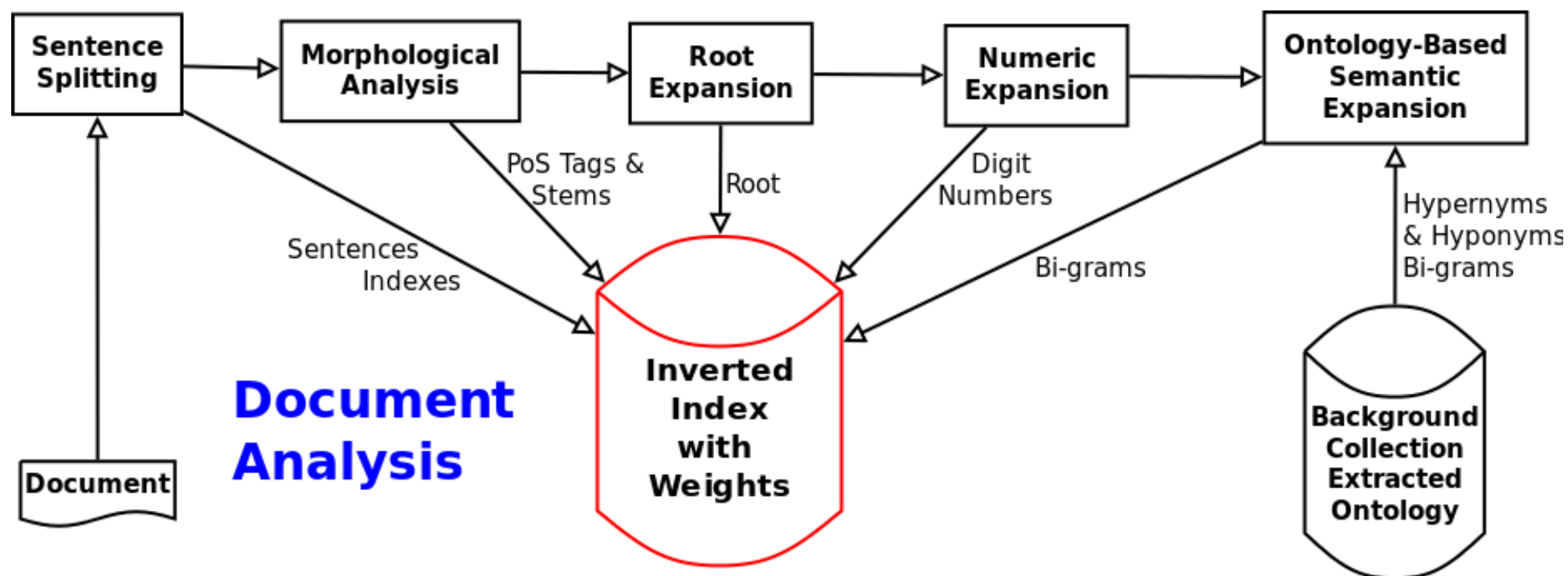
Question Analysis Module
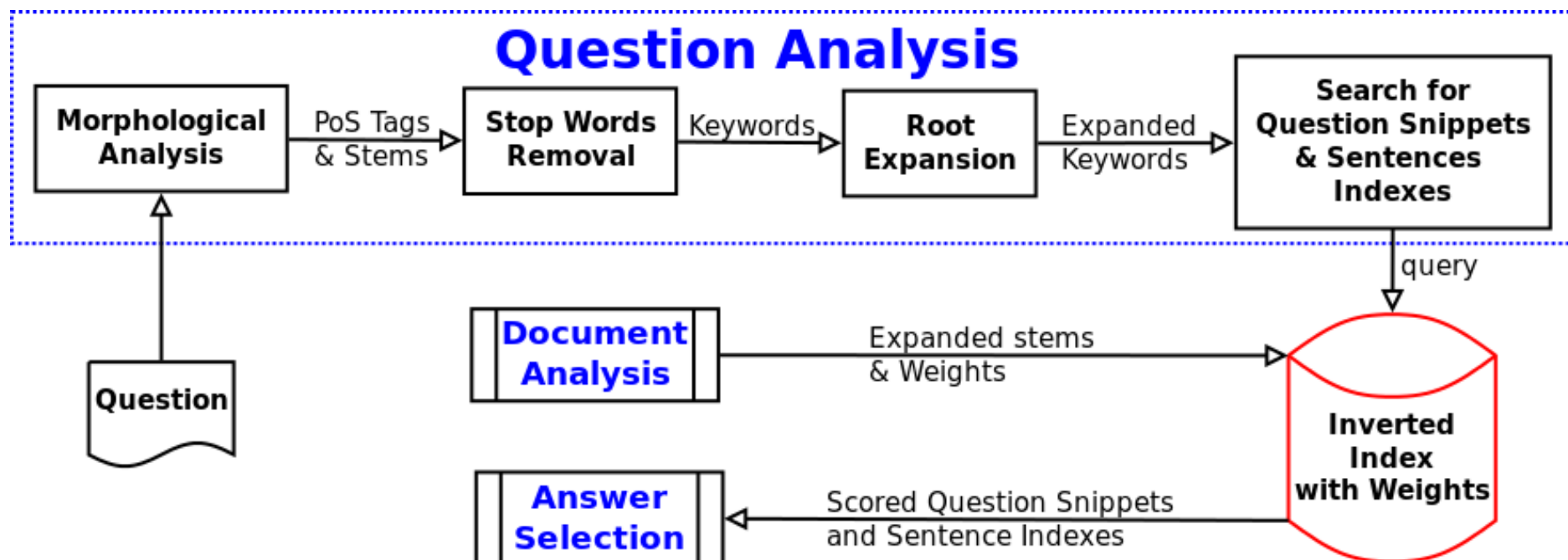
Answer Selection Module
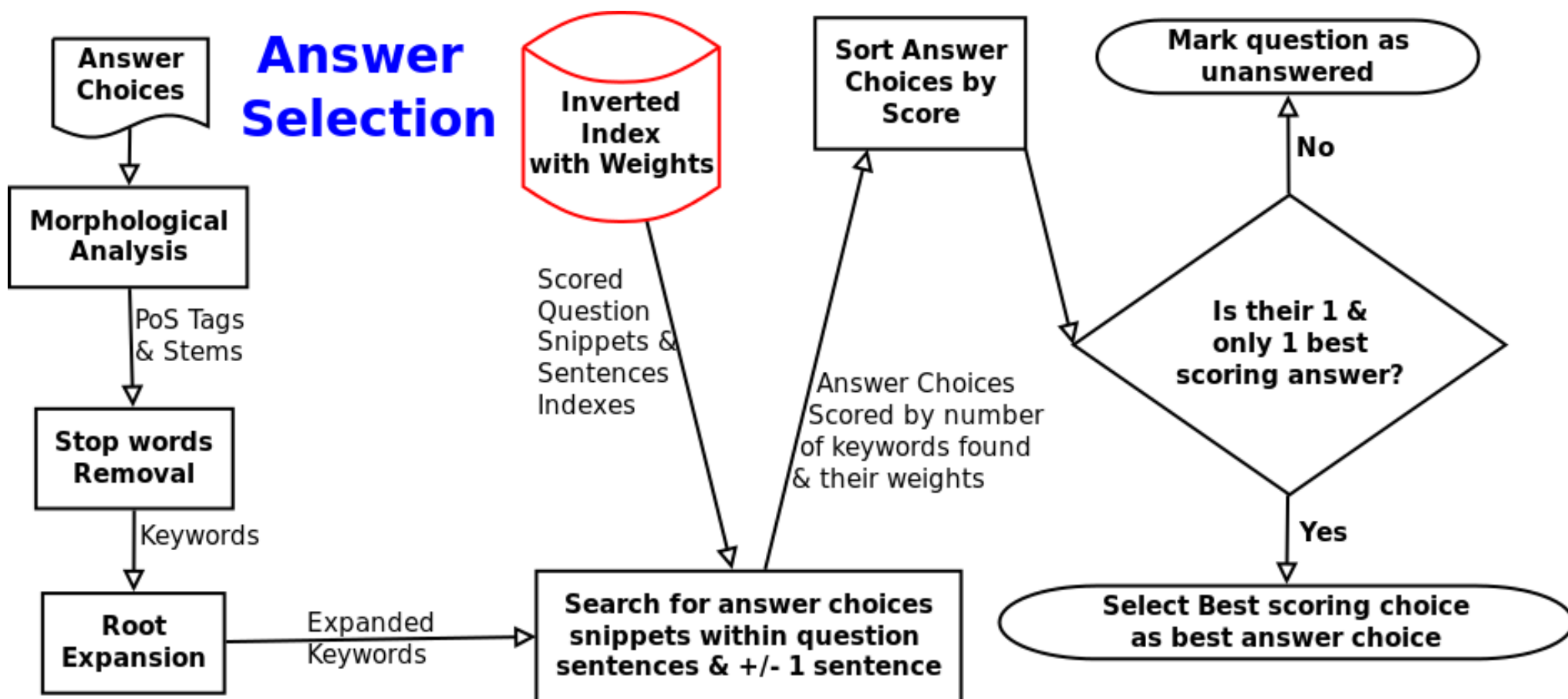
ALQASIM 2.0 [continued]

A basic ontology of hypernyms and their hyponyms generated automatically from the background collection
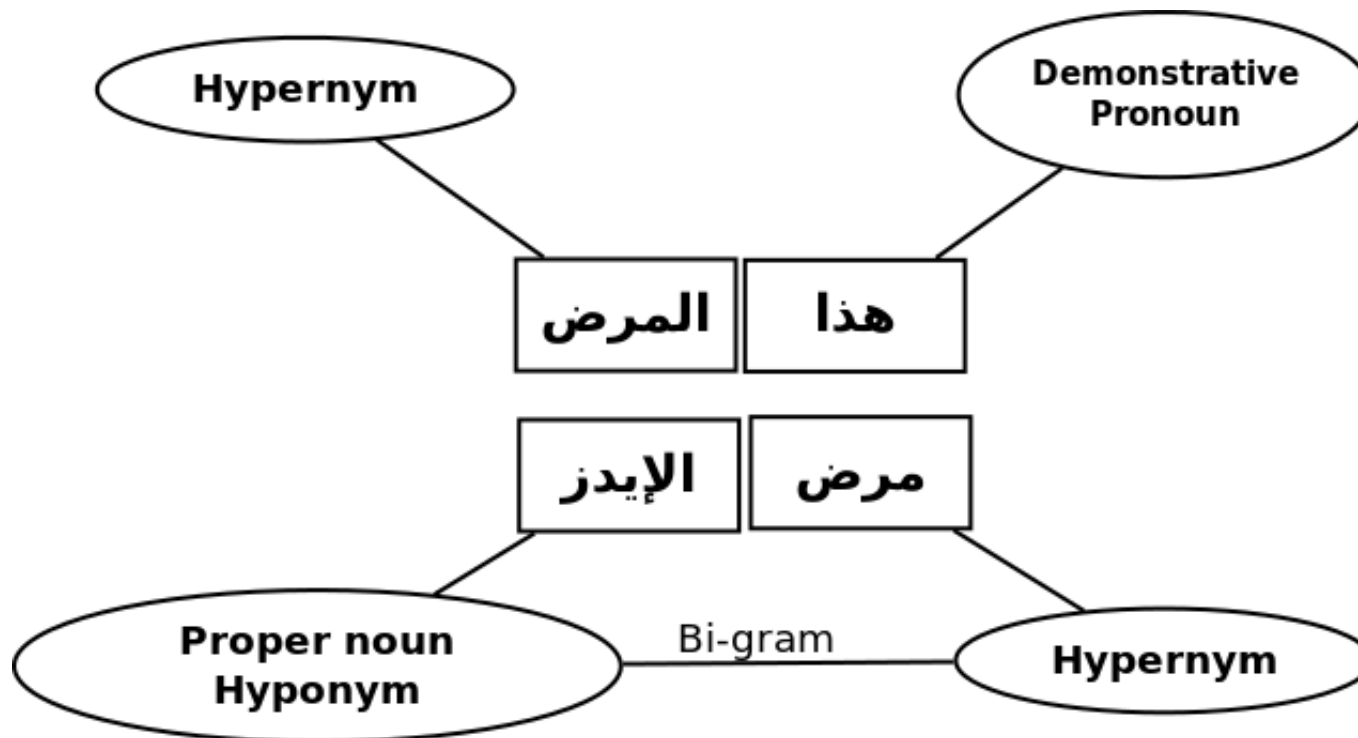
Example 1: Sentence Splitting

Example 2: Ontology Expansion

Example 3: Root Expansion

Evaluation

Description	Correct	Unanswered	Wrong	Accuracy	C@1
ALQASIM 1.0	49	30	81	0.31	0.36
Run (1) Baseline	46	51	63	0.29	0.38
Run (2) Baseline + AWN Semantic Expansion	45	46	69	0.28	0.36
Run (3) Baseline + Bg Ontology Expansion	51	41	68	0.32	0.40
Run (4) Baseline + Bg Ontology Expansion + Root Expansion	57	29	74	**0.36**	**0.42**

Evaluation [continued]
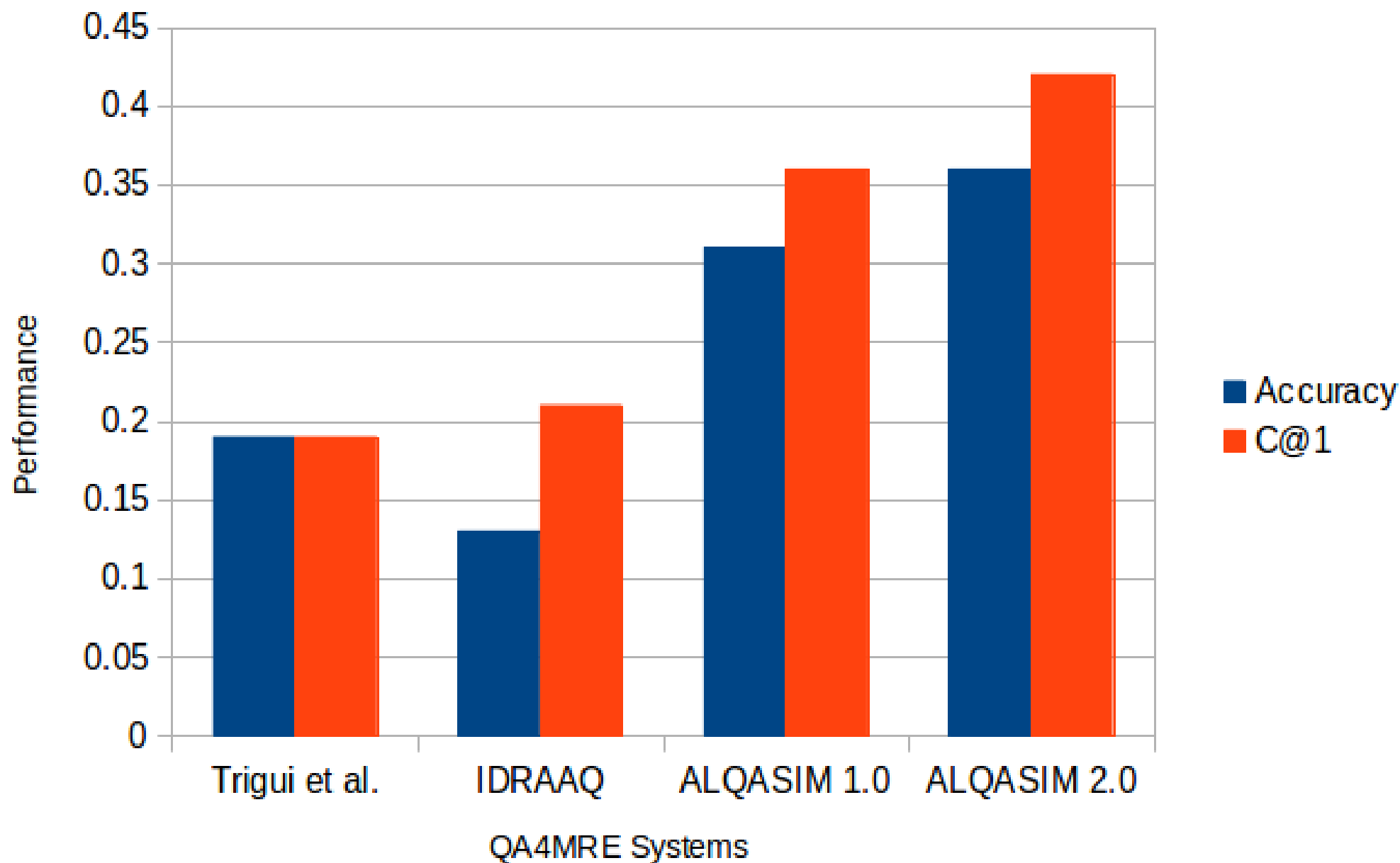
Evaluation [continued]

Root Stemmer	Accuracy	C@1
ALQASIM 2.0 without root expansion	0.32	0.4
ISRI	**0.36**	**0.42**
Khoja	0.32	0.4
Tashaphyne	0.31	0.36

Discussion

- **Sentence** splitting helps identify the correct and the incorrect answers

- **AWN** expansion degrades performance by 1 to 2% due to its generic nature

- **Root** expansion helps answer selection especially with highly derivational texts

- An automatically generated **ontology** specifically created for the test topic improves performance and can be used to boost some question patterns

Root Stemmers

- Khoja vs ISRI on 15163 words from the test-set.

	Khoja		ISRI	
Possible root (a token that is not equal to the light stem or the original word)	36	0.23%	11324	74.68%
Light stem (already generated by MADA)	3647	24.05%	3204	21.13%
Original Word	11480	75.71%	635	4.18%

Tashaphyne Root	Original Words	English Translation
ء	"ماء"، "سوءا"، "سواء"، "وباء"	"water", "worse", "alike", "epidemic"
آا	"آلات"، "آفات"	"machinery", "pests"
أ	"أي"، "أتى"	"any", "came"

Future Work

- Automatically building better **ontologies**

- **Anaphora** resolution

- **Semantic** parsing and semantic role labeling

- Applying Arabic specific **rule-based** techniques

Published Papers

- Ezzeldin, A. M., Kholief, M. H., & El-Sonbaty, Y. (September **2013**). ALQASIM: Arabic language question answer selection in machines. In Information Access Evaluation. Multilinguality, Multimodality, and Visualization (pp. 100-103). Springer Berlin Heidelberg.

- Ezzeldin, A. M., El-Sonbaty, Y., & Kholief, M. H. (October **2014**). Exploring the Effects of Root Expansion, Sentence Splitting and Ontology on Arabic Answer Selection. In Proceedings of the 11th International Workshop on Natural Language Processing and Cognitive Science (NLPCS 2014), Venice, Italy.

Thank you