

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**

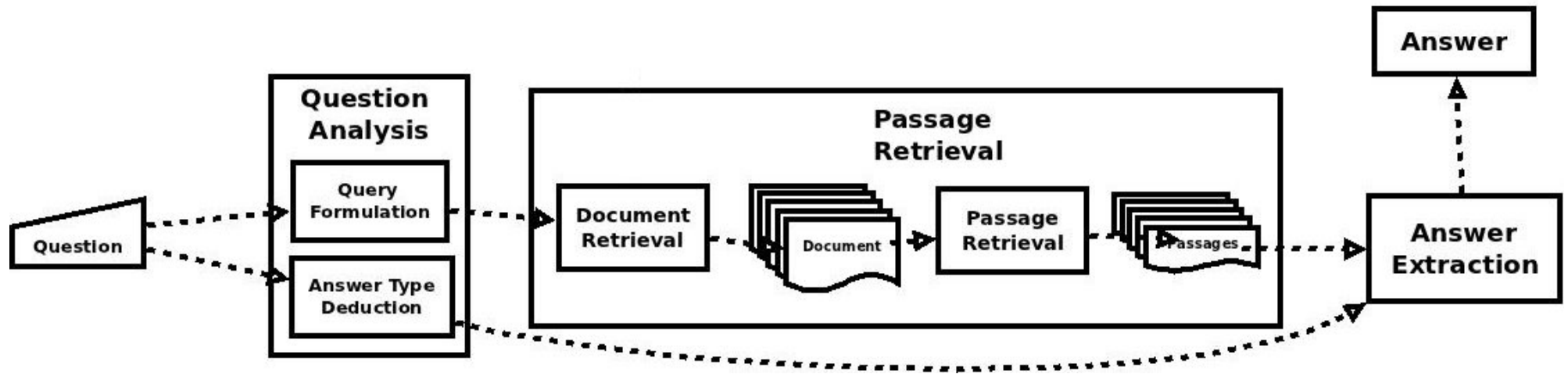
Outline

- 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 from **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

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

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

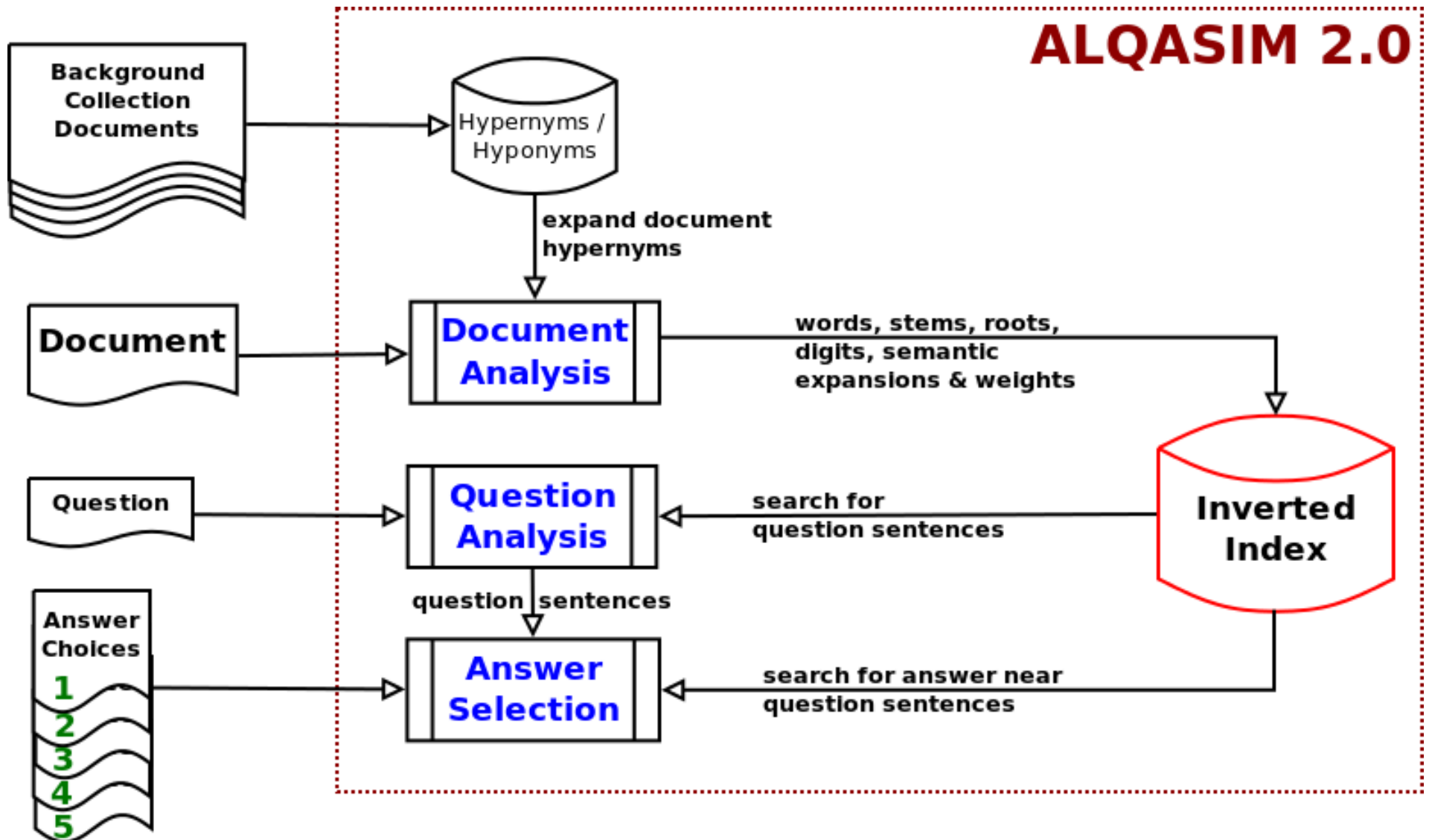
Related Works [continued]

- 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.

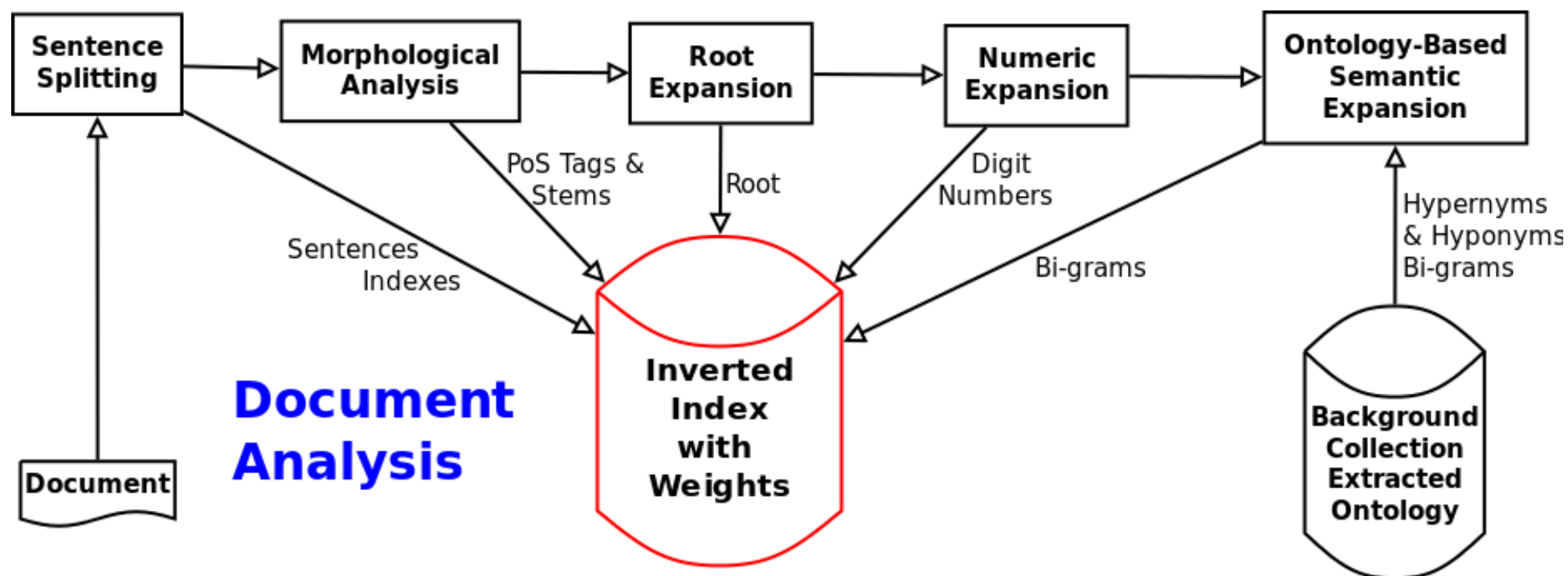
Tools and resources

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

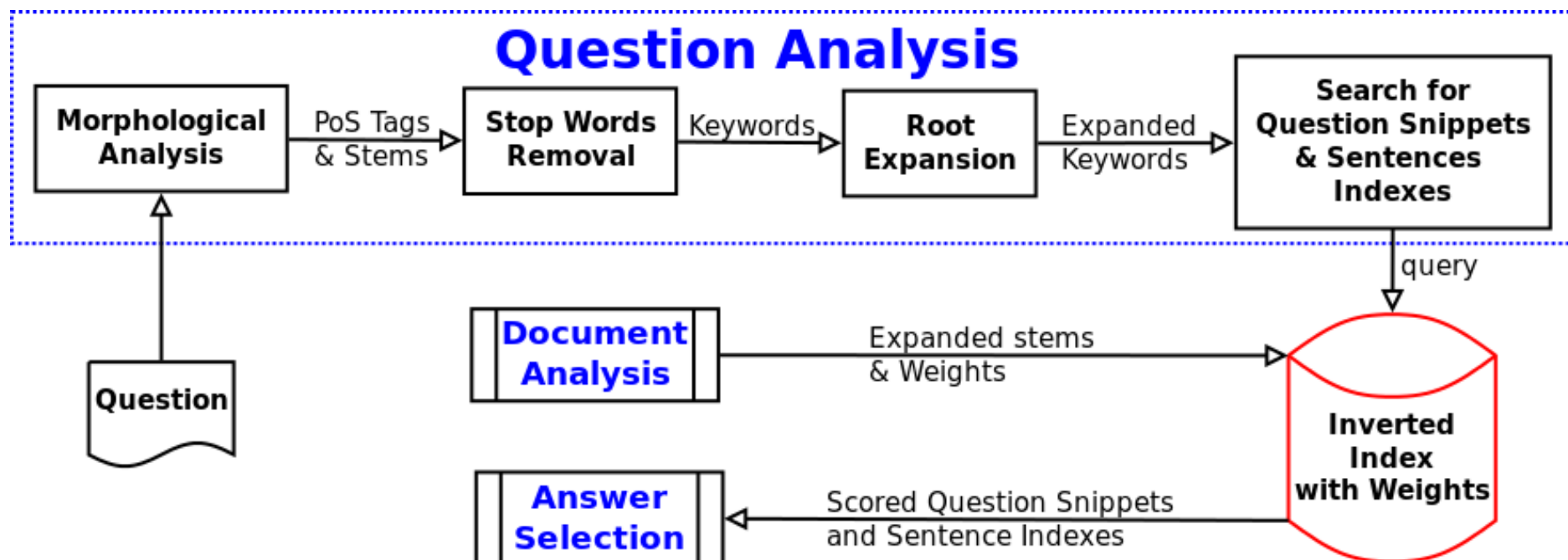
High Level Architecture



Document Analysis Module

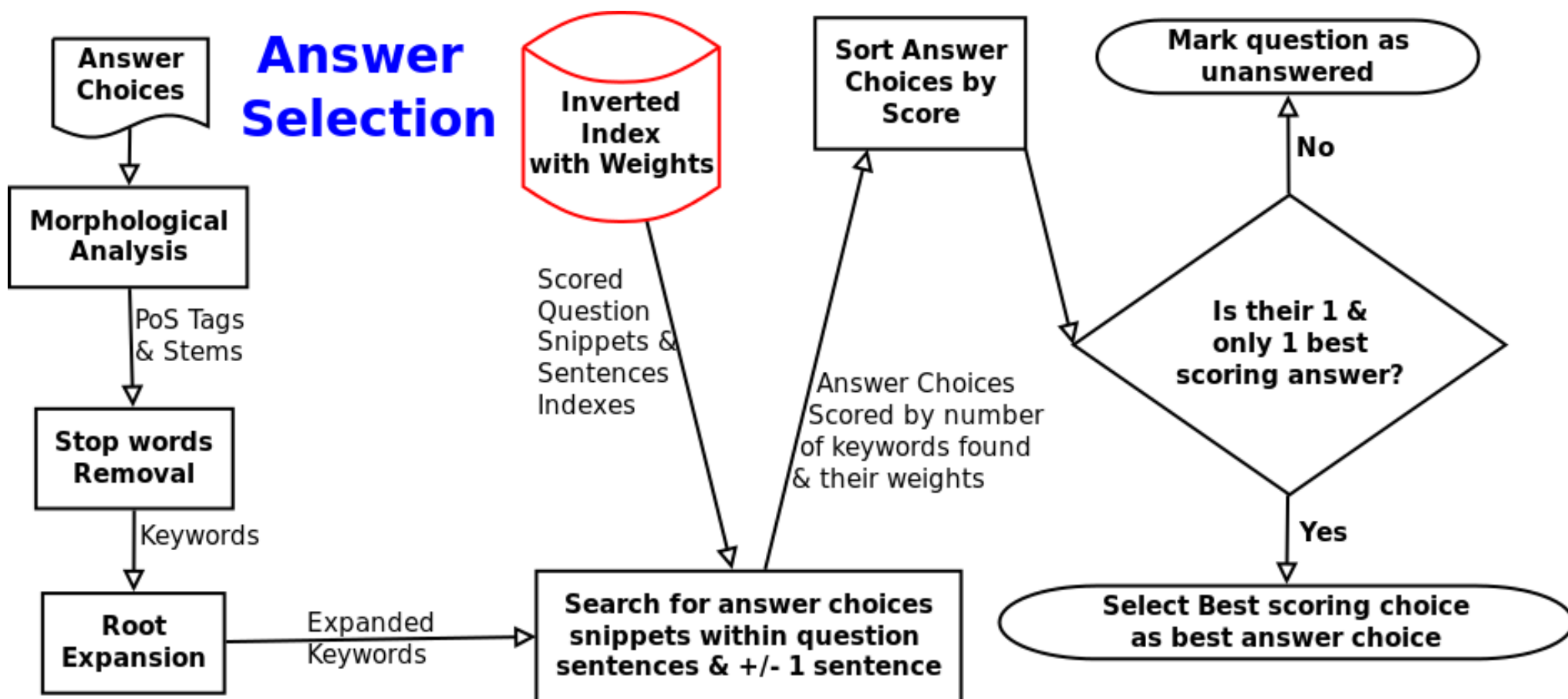


Question Analysis Module



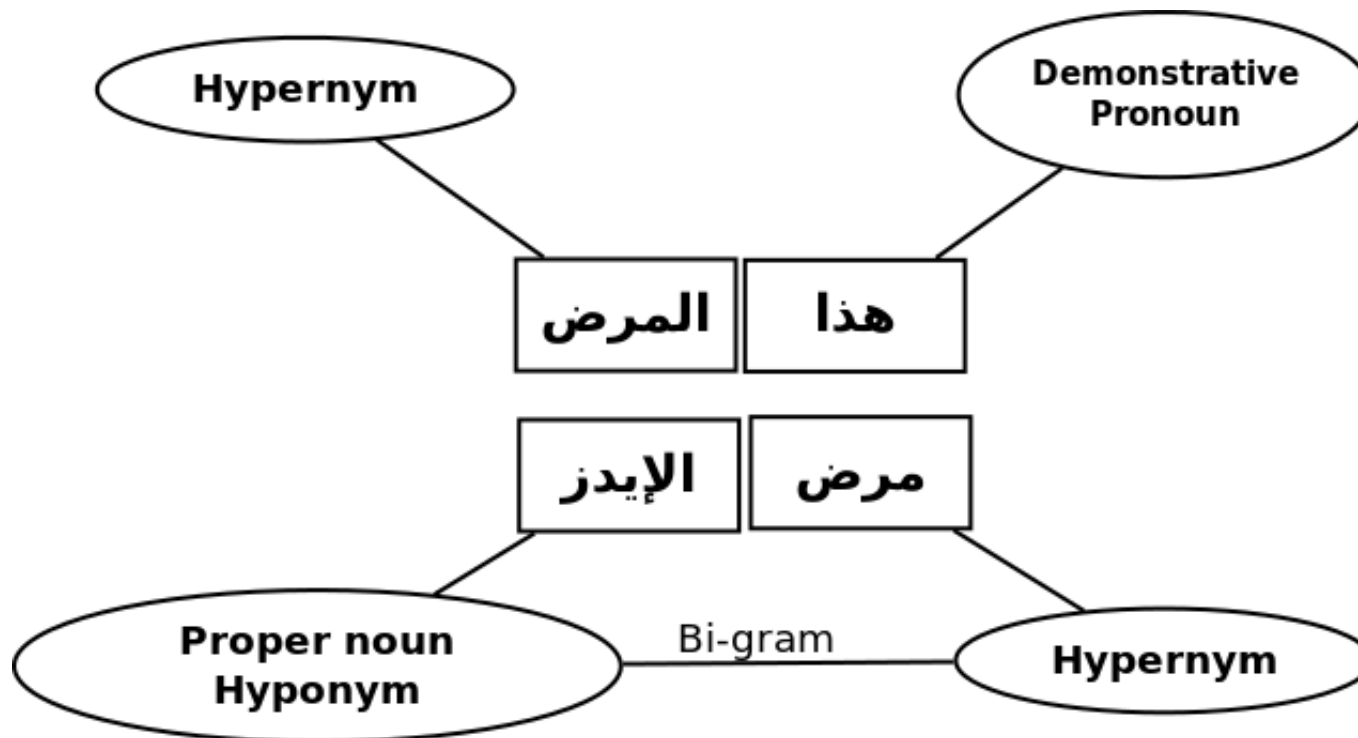
Answer Selection Module

Answer Selection



ALQASIM 2.0 [continued]

A basic ontology of hypernyms and their hyponyms generated automatically from the background collection



Example 1: Sentence Splitting

ALQASIM 2.0

تغيير القوانين والتقاليد وتبسيط الضوء على امور حيوية مرتبطة بحياة الارض وتوزيع الموارد في بلد فقير. وفي الواقع، فإن أي استراتيجية منع ناجحة يجب عليها أن تقوم على دعم الحقوق الاقتصادية والاجتماعية للمرأة. ولكن استراتيجية إدارة بوش السائدة والقاصرة في أن واحد: الامتناع (عن ممارسة الجنس)، الإخلاص، استخدام الواقي الذكري. لا يعتبر خيار الامتناع عن ممارسة الجنس متاحاً بالنسبة للمرأة في حالة الاغتصاب أو الإكراه على الممارسة. أما بالنسبة للإخلاص فغير مجزي في حالة تزوج الرجل بأكثر من واحدة (ويعتبر الزواج بالنسبة للنساء في أفريقيا عامل للإصابة بالمرض). وبالنسبة للواقي الذكري - ويعتبر الحل الأخير بالنسبة لإدارة بوش في معركة مواجهة الإيدز - يعتمد على رغبة الرجل في استخدامه وعلى رغبة الطرفين على تحديد النسل. ومع وضع متاعب المواجهة على عاتق السلوك الفردي، فإن الاستراتيجية المذكورة تتيج لصانعي القرار الفرصة لتجاهل مشاكل الفقر وعدم المساواة، التجاهل الذي يوفر أرضاً خصبة للإيدز. وبينما ينتهي العام 2005، فمن الواضح أن استراتيجية الأمم المتحدة، مبادرة "3 في 5"، لاتاحة مضادات فيروسية لثلاث ملايين نسمة مع نهاية العام ستكون فاشلة بنسبة الثلثين. ففي أفريقيا، لا يستطيع 9 من كل 10 من مرضى الإيدز الوصول لهذه الأدوية، هذه الأدوية المتاحة لكافة في الدول الغنية. والسبب؟ غياب الإرادة السياسية وارتفاع أسعار هذه الأدوية. وصول الكافة للعلاج هو هدف يمكن الوصول إليه، ولكنه يحتاج قرار من الولايات المتحدة والاتحاد الأفريقي في لقاء منظمة التجارة العالمية لهذا الشهر لاحترام حق الدول الفقيرة في استيراد الأنواع الرخيصة لأدوية الإيدز. جرت برامج فعالة تهدف إلى منع ومعالجة الإيدز معاً في بلاد مثل أوغندا، تنزانيا، تايلاند ودول أخرى. ولكن بدلاً من دعم أحسن هذه البرامج، أعطت إدارة بوش السياسات المرتبطة بالإيدز في أيدي المسيحيين الأصوليين (الذين ضغطوا من أجل تطبيق سياسة الامتناع والإخلاص والعازل الذكري المتفق مع أيديولوجيتهم) بالإضافة إلى جماعات الضغط المرتبطة بشركات الأدوية (الساعين إلى تبني أجندة التي تخدم مصالحهم الخاصة على حساب الصحة العامة).

Question Snippets
Sentence: 29 Score: 794.00

أفريقيا
الأدوية
دواء
دوي
رضي
مرضى

Answer [2] Snippets
Sentence: 30 Score: 402.00

أسعار
ارتفاع
رفع
سعر

Topic: [AIDS] Document: [1] Question: [1]

الحصول على مضادات الفيروسات الرجعية الأدوية للمرضى في أفريقيا؟

توافر العقاقير المضادة للفيروس في البلدان الغنية [1] 178.00

ارتفاع أسعار الأدوية [2] 402.00 - C - A

إلغاء الديون الخارجية للحكومات الأفريقية [3] 56.00

أرباح شركات الأدوية [4] 200.00

عدم وجود خطط سياسية [5] 119.00

Example 2: Ontology Expansion

ALQASIM 2.0

الاجتماعي في افريقيا مثل ريبيكا، يزداد عدد الذين يعرفون ان افريقيا جنوب الصحراء هي الان المركز الأساسي لوباء الإيدز: فافريقيا موطن 3 من كل 4 في العالم يموتون بسبب المرض، وهي القارة التي بها ثلثي حاملي الفيروس في العالم (أكثر من 25 مليون نسمة). ولكن لا يعرف الجميع أن معظم هؤلاء المرضى هم من النساء، وأن حديثي السن منهن يزيد معدل إصابتهن بمعدل من 3 إلى 4 أضعاف الرجال من نفس الشريحة السنية. وللكثيرين فإن هذه المعلومات يصعب تأثيرها عبر شبكة من الأنماط التي تجعل هذا الكارثة الإنسانية هي ظروف الحياة الطبيعية في أفريقيا. وبينما نشأ الإيدز - مثل الكوارث الطبيعية الأخرى لهذه السنة - من الطبيعة، فإن ما يحدثه من آثار مدمرة يجعله كارثة صنعها الإنسان. فيجب أخذ التالي في الاعتبار: منذ الثمانينات عندما بدأ ظهور الإيدز، طالبت الولايات المتحدة بسياسات اقتصادية قاسية في الدول الفقيرة. ففي أفريقيا، قطعت هذه السياسات ميزانيات الصحة إلى النصف بينما كانت أنظمة الصحة العامة تحتاج هذا الدعم لمواجهة الإيدز. واليوم، فإن هذا الوباء هو العائق الأكبر الوحيد تجاه التنمية الاقتصادية في أفريقيا. ولزيادة الأرباح الفاحشة لشركات الأدوية الأمريكية، فإن إدارة بوش منعت بيع الأدوية الرخيصة الثمن التي كان يمكنها إنقاذ حياة الملايين في الدول الغنية. تعرضت النساء لخطر أكبر تجاه الإيدز لعدم إعطائهن الحق في الامتناع عن ممارسة الجنس أو الإصرار على استعمال الواقي الذكري. وبينما يعيش معظم في ظروف فقيرة تتعرض النساء الأكثر فقراً للفيروس أكثر من غيرهن، بل ويزيد احتمالية نشاط الفيروس عندهن. تعتبر عوامل الإيدز، والسياسات الاقتصادية الغير عادلة، وعدم المساواة بين النساء والرجال مثل أزمات، فمواجهة أي ضلع من هذا المثلث يستدعي مواجهة الضلعين الآخرين. ولكن في أحيانا كثيرة فإن برامج الصحة العامة، والسياسات الحكومية - وحتى حملات النشاط تركز في جزء واحد مما يؤدي إلى تفويت نقط مهمة في مواجهة هذا المثلث مما يقلل القدرة على التغيير. من أسباب هذه النظرة القاصرة هي أنها تتيح لصناع القرار التنصل من

Question Snippets
Sentence: 10 Score: 780.00

أكبر
اقتصادي
اقتصادية
الاقتصادية
التنمية
تنمية
نمي

Answer [2] Snippets
Sentence: 9 Score: 601.00

إيدز
الأيدز

Topic: [AIDS] Document: [1] Question: [7]

ما هي أكبر عقبة أمام التنمية الاقتصادية في "القارة السوداء"؟

ABC استراتيجية 0.00 [1]

الإيدز 601.00 - C - A [2]

مادر 0.00 [3]

صندوق النقد الدولي 0.00 [4]

الكوارث الطبيعية 0.00 [5]

Example 3: Root Expansion

ALQASIM 2.0

لتخطيط الطاقة، والهدف من هذا هو تقليل المخاطر البيئية للمستثمرين، و تحفيز استثمارات الطاقة الكهرومائية. برامج أخرى للطاقة في عام 2002، اقترحت البرازيل مبادرة الطاقة البرازيلية، والتي تضع أهدافا لزيادة استخدام الطاقة المتجددة في جميع أنحاء أمريكا اللاتينية. كان الهدف المنشود أن يتم توفير عشرة في المائة من احتياجات الطاقة الموردة من مصادر الطاقة المتجددة بحلول عام 2010. حظيت المبادرة بدعم قوي من وزراء البيئة من دول أمريكا اللاتينية ودول البحر الكاريبي في اجتماع عقد في مايو 2002، كما أثرت أيضا في المفاوضات حول الطاقات المتجددة خلال القمة العالمية للتنمية المستدامة. دعت الوثيقة الختامية للقمة الدول إلى "العمل على وجه الاستعجال إلى زيادة كبيرة" في الحصة العالمية من الطاقة المتجددة. كما حثت القمة الدول على تحويل الدعم الحكومي من الوقود الأحفوري والنووي إلى مصادر الطاقة النظيفة، من أجل تشجيع النمو في قطاعات الطاقة المتجددة. أن البرازيل وإحدى من البلدان القليلة في العالم التي تستخدم الفحم في صناعة الحديد والصلب. ولأن الفحم يأتي عن طريق محاصيل متجددة، فقد أدى ذلك إلى تقليل انبعاثات غاز ثاني أكسيد الكربون في القطاع الصناعي بخمسين مليون طن بين عامي 1990 و2000. بدأت البرازيل مؤخرا البرنامج الوطني للوقود الحيوي (البيوديزل)، الذي يهدف إلى زيادة متنامية في نسبة الوقود الحيوي في وقود الديزل المستخدم في جميع أنحاء البلاد. إن هدف البرنامج هو إضافة ما لا يقل عن ثلاثة في المائة من البيوديزل إلى وقود الديزل الأحفوري بحلول عام 2008، وخمسة في المائة بحلول عام 2012. حاليا، لا يتم استخدام البيوديزل في البرازيل. يمكن إنتاج البيوديزل من عدة أنواع فول ونخيل تنمو بسهولة في البرازيل، مثل زيت النخيل والخروع وفول الصويا وعباد الشمس والفول السوداني وغيرهم. ويمكن أيضا أن ينتج البيوديزل من الدهون الحيوانية والصرف الصحي والزيت النباتي المستخدم. يهدف البرنامج الوطني للوقود الحيوي إلى تشجيع البيوديزل المصنوع من بذور

Question Snippets

Sentence: 66 Score: 1486.00

البرازيل
البرنامج
الحيوي
برازيل
برنامج
حيوي
حيي
نوع

Answer [4] Snippets

Sentence: 66 Score: 500.00

الديزل
ديزل
زبد

Topic: [Climate Change] Document: [6] Question: [8]

ما هو الهدف من البرنامج القومي للوقود الحيوي في البرازيل؟

الجمع بين وقود الديزل الحيوي والإيثانول

زيادة إنتاج وقود الديزل الحيوي من نبات الذرة

إنتاج وقود الديزل الحيوي من الدهون الحيوانية

لتشمل المزيد من وقود الديزل الحيوي في وقود الديزل العادي

لزيادة واردات وقود الديزل الحيوي

☐ 300.00

☐ 399.00

☐ 198.00

☒ 500.00 - C - A

☐ 401.00

[1]

[2]

[3]

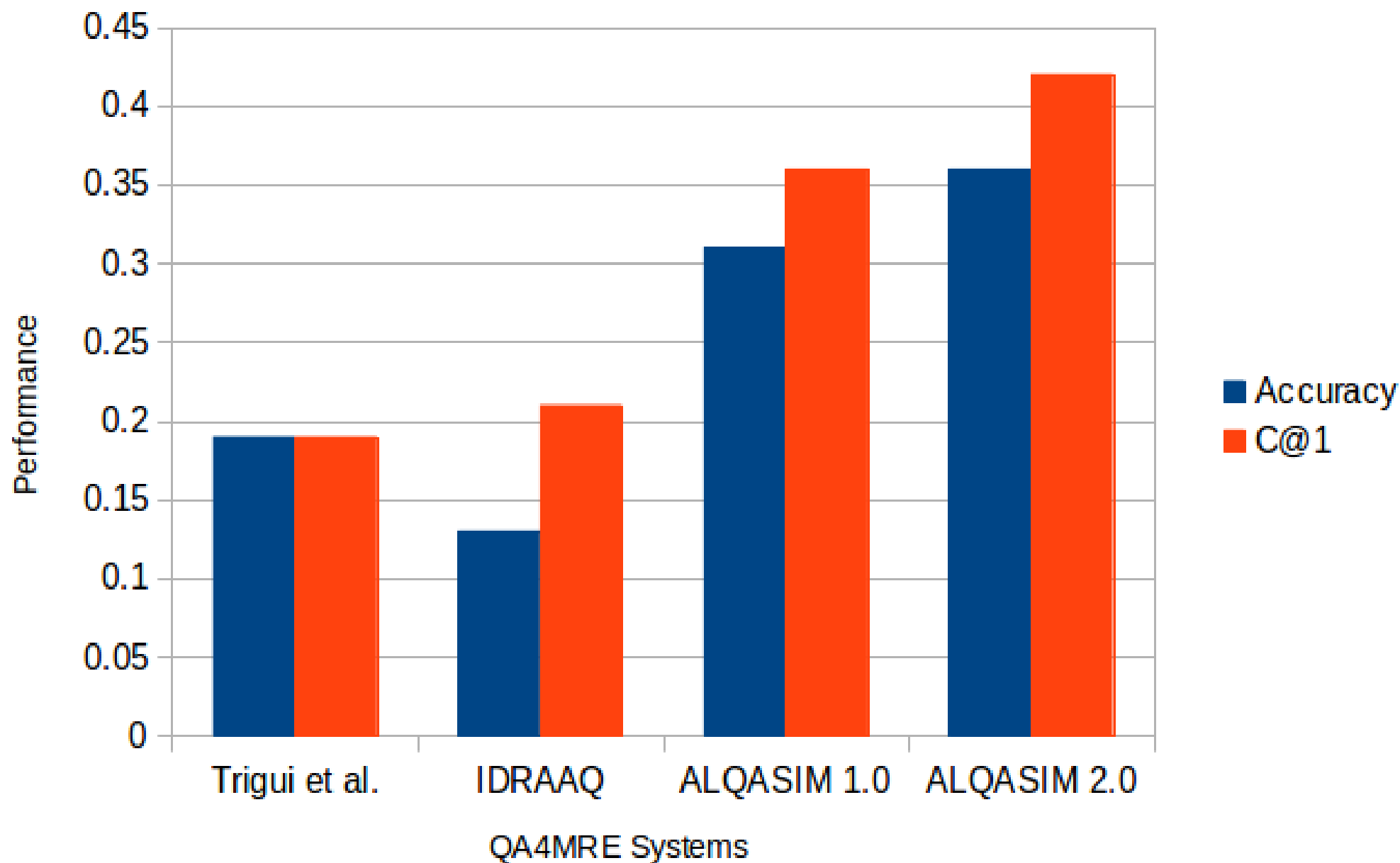
[4]

[5]

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"
آ	"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