



*twas*

# ANSWER EXTRACTION FROM AN ACADEMIC QUESTION ANSWERING SITE

**Bolanle Ojokoh**

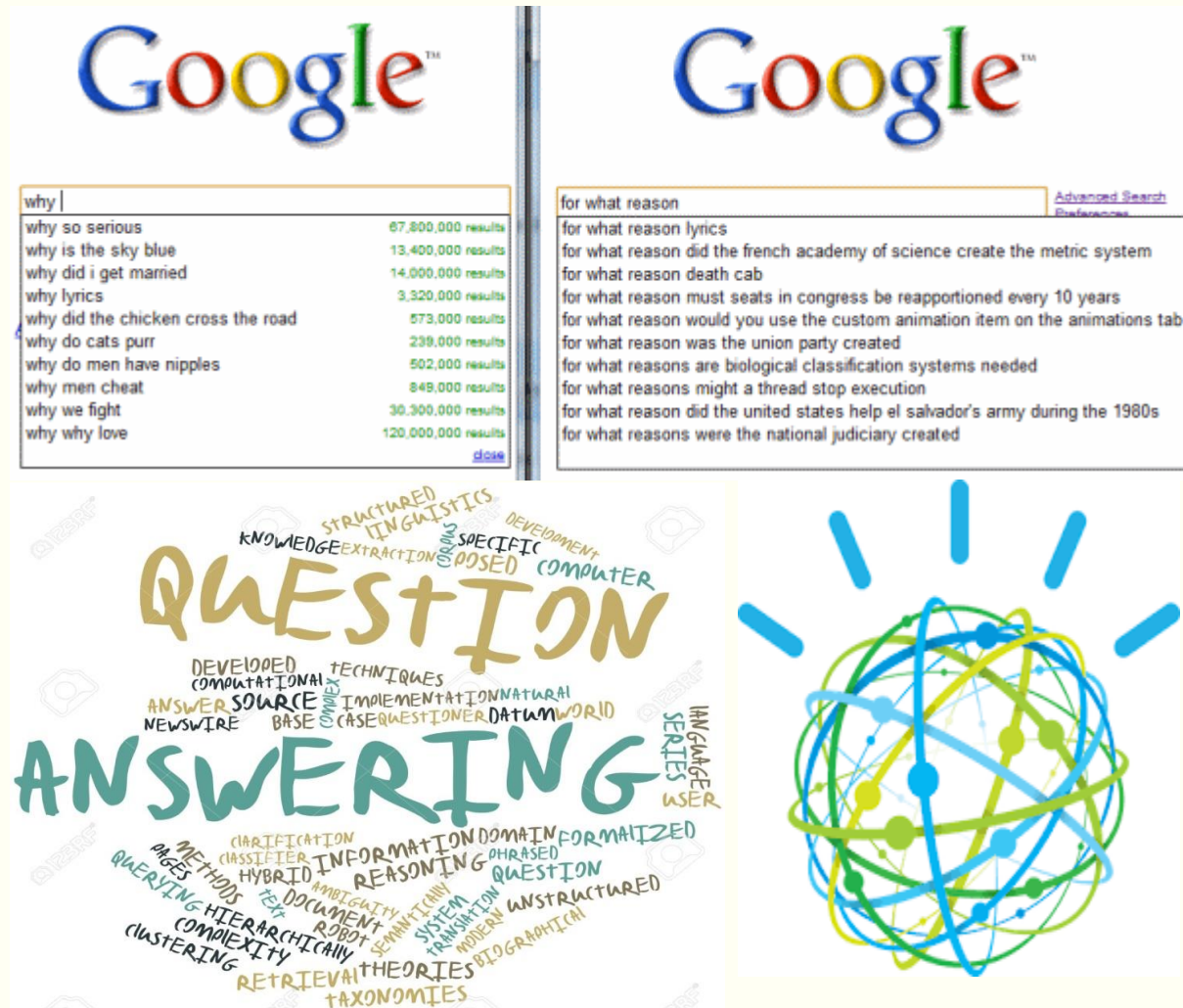
**Federal University of Technology, Akure, Nigeria.**



**1<sup>st</sup> International  
Conference of  
TWAS Young  
Affiliates Network**

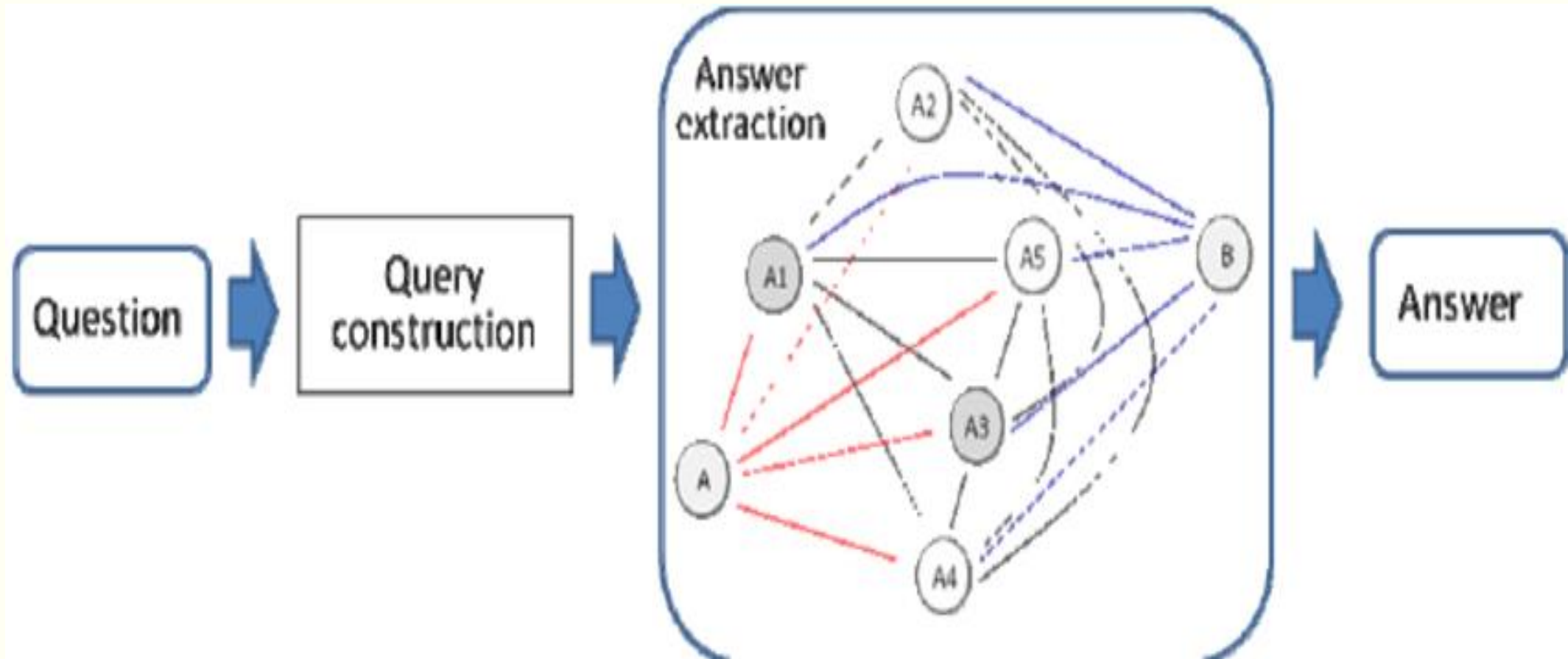
# What is Question Answering?

- QA is an interactive human computer process that involves providing answers to meet a user information need presented in the form of a query



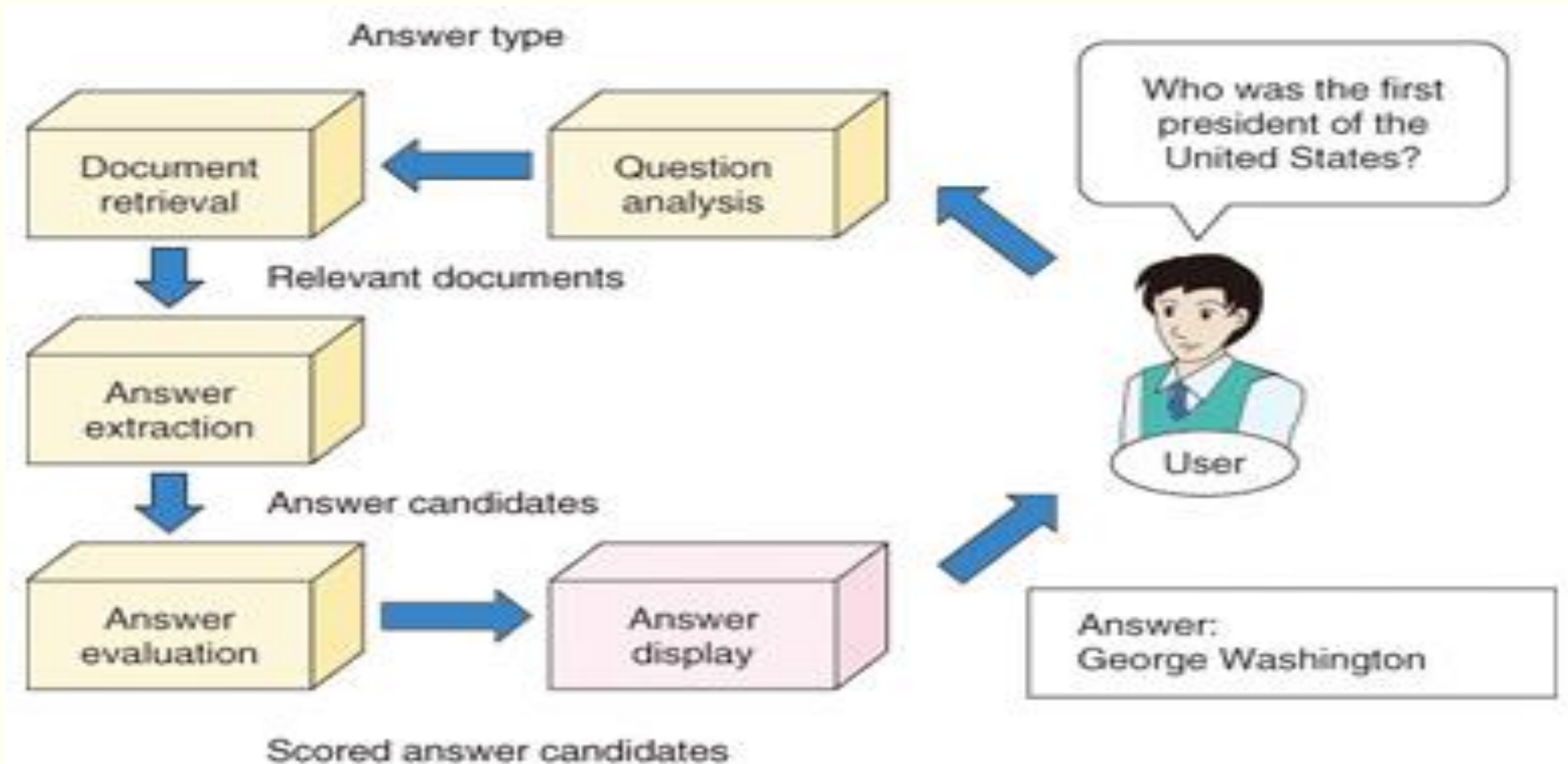
# Question Answering Process

---



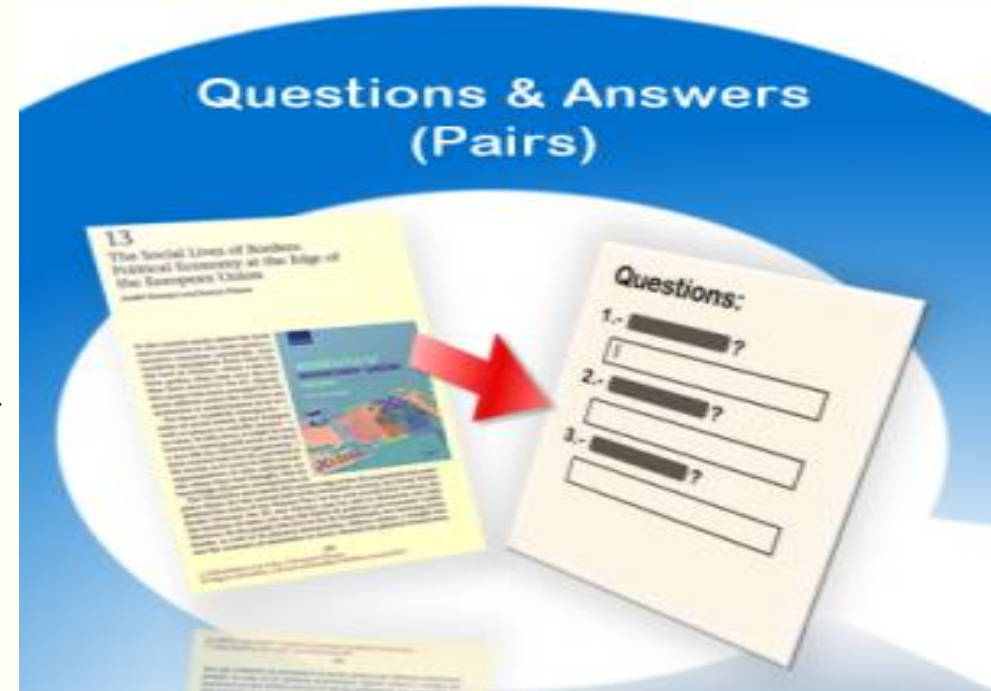
**retrieving** relevant documents, data, or knowledge from selected sources, **extracting**, **qualifying** and **prioritizing** available answers from these sources

# Question Answering Process



# Goal – Extracting Quality QA pairs from Online Forums

---



# Our Case Study - Researchgate

- Researchgate is a social networking site for scientists and researchers to share papers, ask and answer questions, and find collaborators.

The image shows a ResearchGate profile for Jayendra Kumar and a detailed view of his RG Score. The profile includes a question about ANN hidden layers, a share button, and an answer from Olivier Parisot. The RG Score section shows a score of 38.68, a pie chart of its components, a percentile indicator, and a bar chart showing the score over time.

## ResearchGate

**Jayendra Kumar** 4.60  
National Institute of Technology, Jamshedpur

**How can I decide number of neuron in hidden layer in ANN for probable best performance in classification problem ?**

Researcher who work in machine learning.

**TOPICS**  
Machine Learning

Dec 12, 2014 · Flag

Share 0 / 0

**ALL ANSWERS (10)**

**Olivier Parisot** 5.63 · Luxembourg Institute of Science and Technology  
Dear Jayendra,  
This page could be useful for you: <http://www.heatonresearch.com/node/707>  
Kind regards,  
Olivier PARISOT

1 / 0 · Dec 12, 2014

### 38.68 RG Score

A new way to measure scientific reputation.  
The RG Score takes all your research and turns it into a source of reputation.

- PUBLICATIONS
- ANSWERS
- QUESTIONS
- FOLLOWERS

**PERCENTILE**  
Your score is higher than 97.5% of ResearchGate members!

0% 50% 100%

#### RG SCORE OVER TIME

Date	RG Score
06-11	0
06-18	0
06-25	0
07-02	0
07-09	0
07-16	0
07-23	0
07-30	22
08-06	42
08-13	42
08-20	38
08-27	36
09-03	35
09-10	35
09-17	35

**How does the RG Score work?**  
Your RG Score is calculated based on how other researchers interact with your content, how often, and who they are. The higher their score, the more yours will increase.

# Our Approach

---

Question Extraction and Classification – Fuzzy Logic+ Part of Speech Tagging+ Naïve Bayes+Cosine Similarity

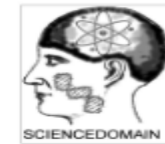


Extensibility



**British Journal of Mathematics & Computer Science**  
22(1): 1-21, 2017; Article no.BJMCS.32541  
ISSN: 2231-0851

**SCIENCEDOMAIN international**  
[www.sciencedomain.org](http://www.sciencedomain.org)



## **Feature-based Model for Extraction and Classification of High Quality Questions in Online Forum**

**Bolanle Ojokoh<sup>1\*</sup>, Tobore Igbe<sup>1</sup> and Ayobami Araoye<sup>1</sup>**

<sup>1</sup>*Department of Computer Science, Federal University of Technology, P.M.B. 704, Akure, Nigeria.*

### *Authors' contributions*

*This work was carried out in collaboration between all authors. Author BO initiated and designed the study. Author TI managed the implementation and experimental study and wrote the first draft of the manuscript. Author AA managed the literature searches and part of the experimental study. All the authors read and approved the final manuscript.*

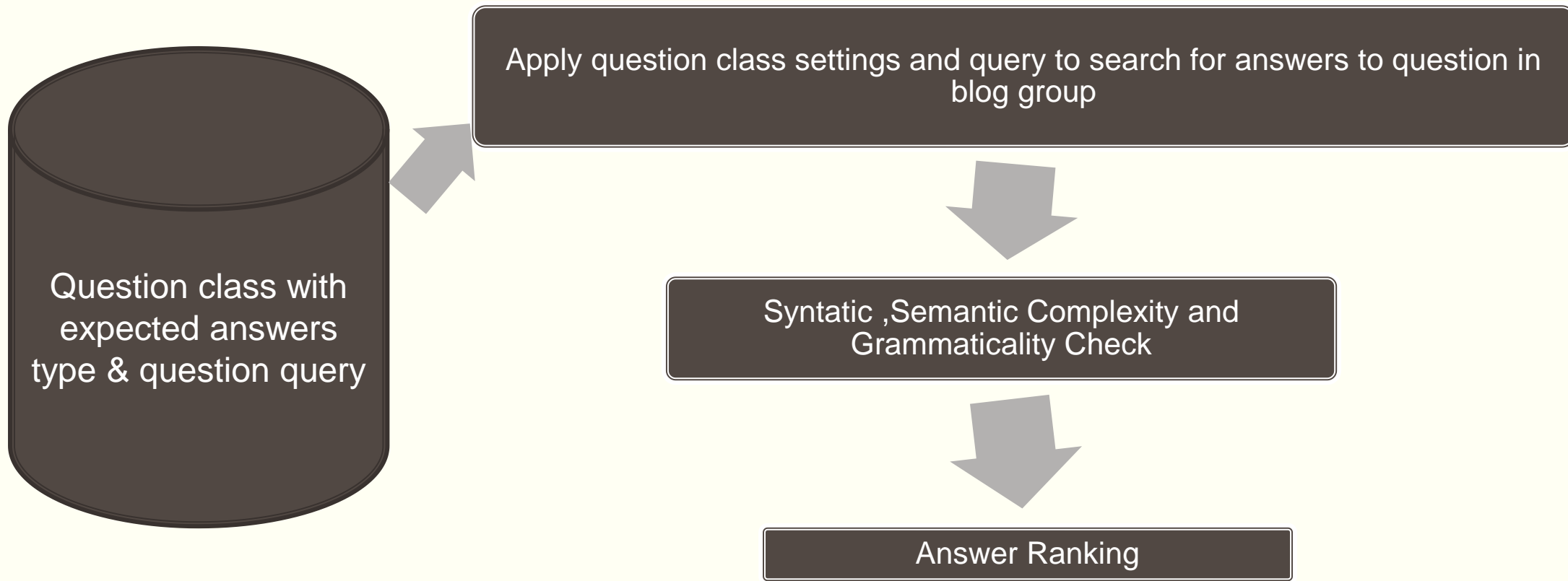
### *Article Information*

DOI: 10.5734/BJMCS.2017.22541

# In this work

---

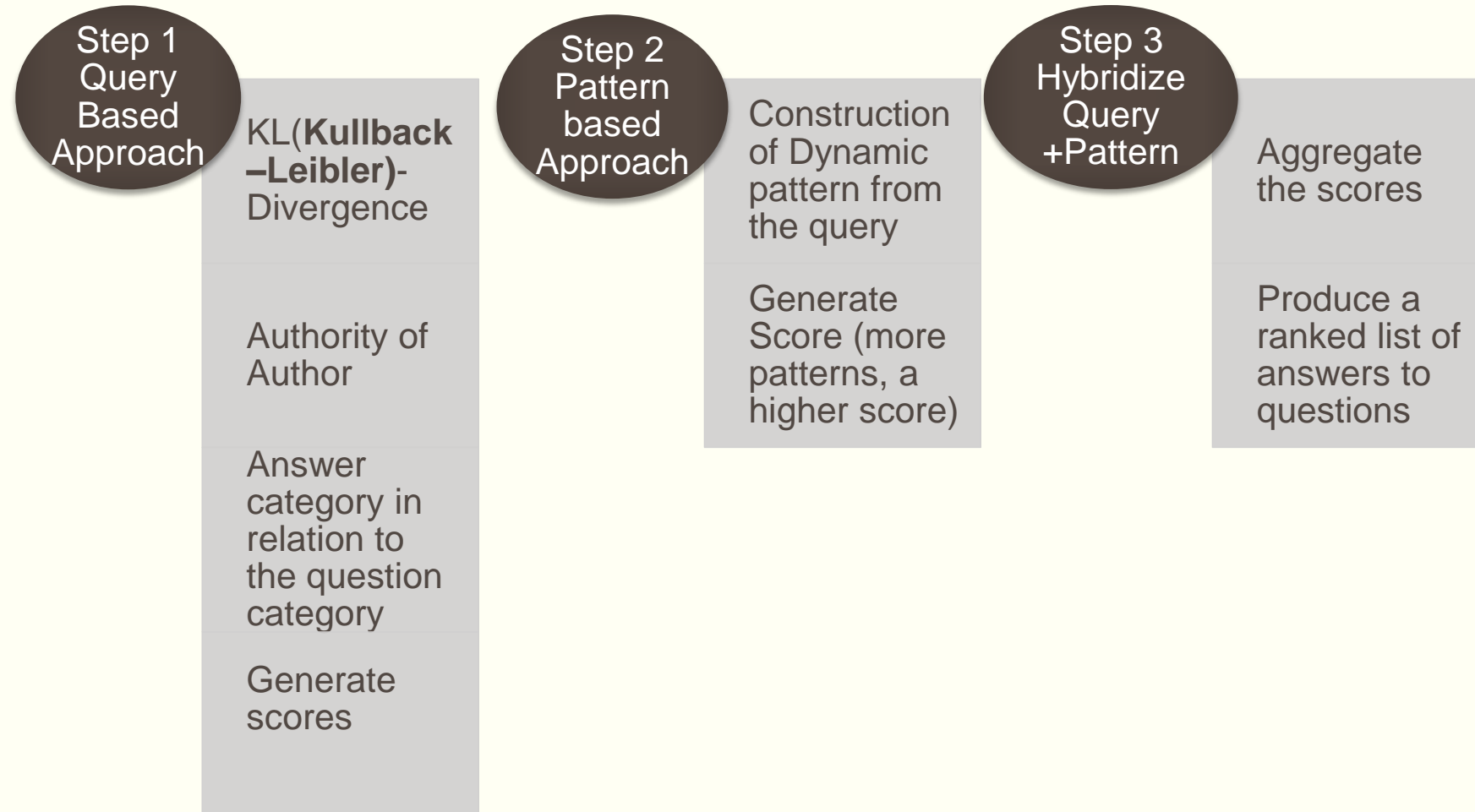
---





# Our Approach

---



## The Idea

---

- More number of replies, more credibility (better answer quality) : having replied many related questions is likely to produce a better answer
- The category of a question has a strong correlation with the likely answer category

# Experimental Results

---

- 1000 dataset files crawled from Researchgate
- Consisting of 24,911 sentences

## Evaluation Results

<b>Answer extraction Methods</b>	<b>Average value</b>			
	<b>Precision</b>	<b>Accuracy</b>	<b>Recall</b>	<b>F-measure</b>
<b>Query Based (QB) method</b>	0.985	1.000	0.988	0.992
<b>Pattern Based(PB) method</b>	0.999	1.000	0.999	0.999
<b>QB+PB method</b>	1.000	1.000	1.000	1.000

## Conclusion and Future Work

---

---

- Isolation of answers for questions with high performance accuracy and precision.
- The combined method produced a better result over the individual methods.
- Test proposed model on other types of dataset generated from other blogs and social media.

# Research Collaborators and Co-Authors

---



Tobore Igbe



Bamidele Afolabi

# Acknowledgments

---

---

- TWAS-COMSTECH for Research grant for Individual Researchers
- TWAS, TWAS-ROSSA and TYAN for travel support
- BAS and Patricia Zancan for hospitality and hosting
- TYAN Executive members for their commitment towards the success of the conference
- Everybody in attendance