

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

دورة "استرجاع المعلومات" باللغة العربية - صيف ٢٠٢١

Information Retrieval – Summer 2021



10. IR Research & Resources

Tamer Elsayed
Qatar University

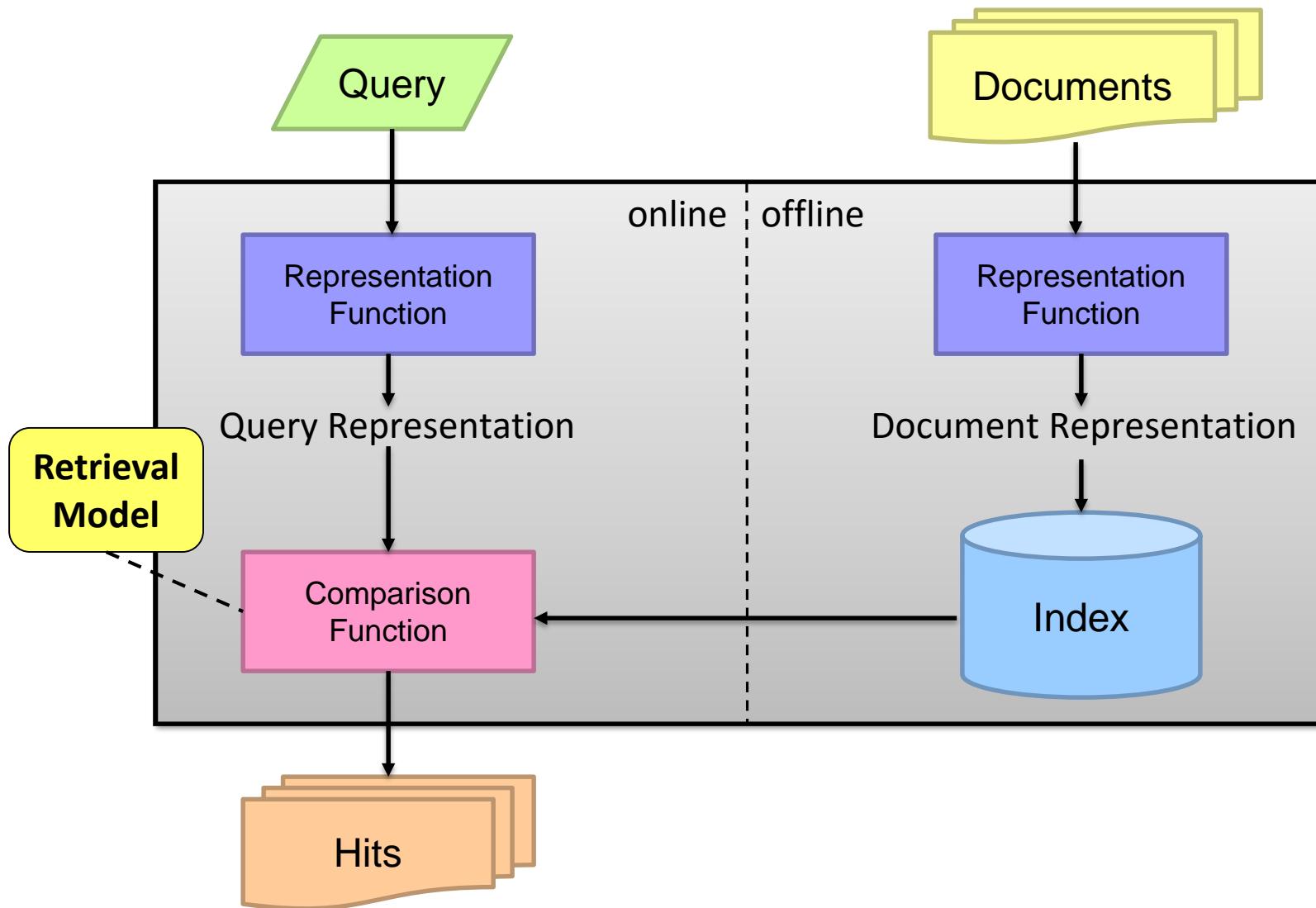
Today's Roadmap

- The BIG picture
- My group's ongoing research
- Open research in IR
- Relevant Jobs
- Resources for Arabic IR
- Project ideas for grad students
- Where can you go from here?

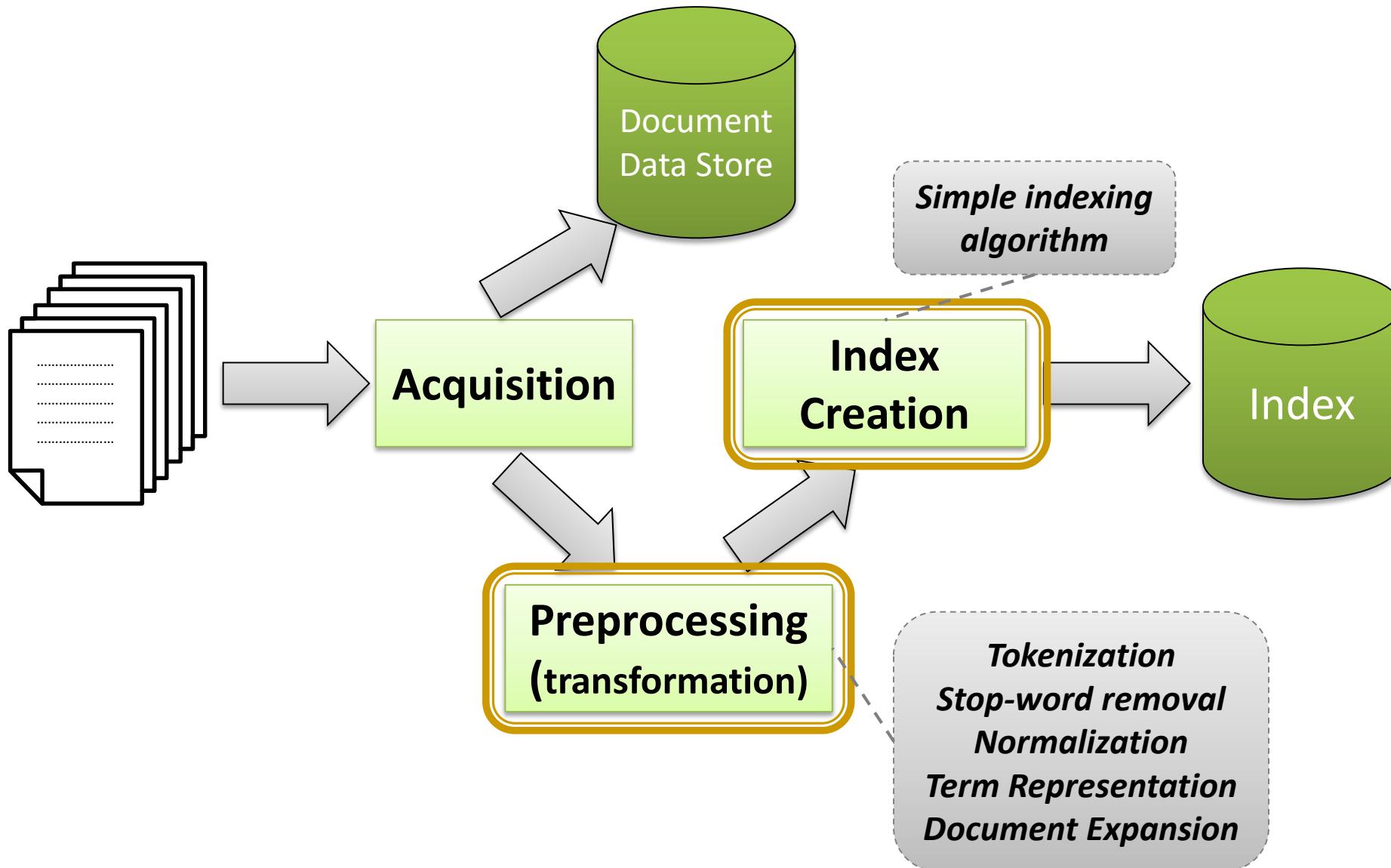


**THE
BIG
PICTURE**

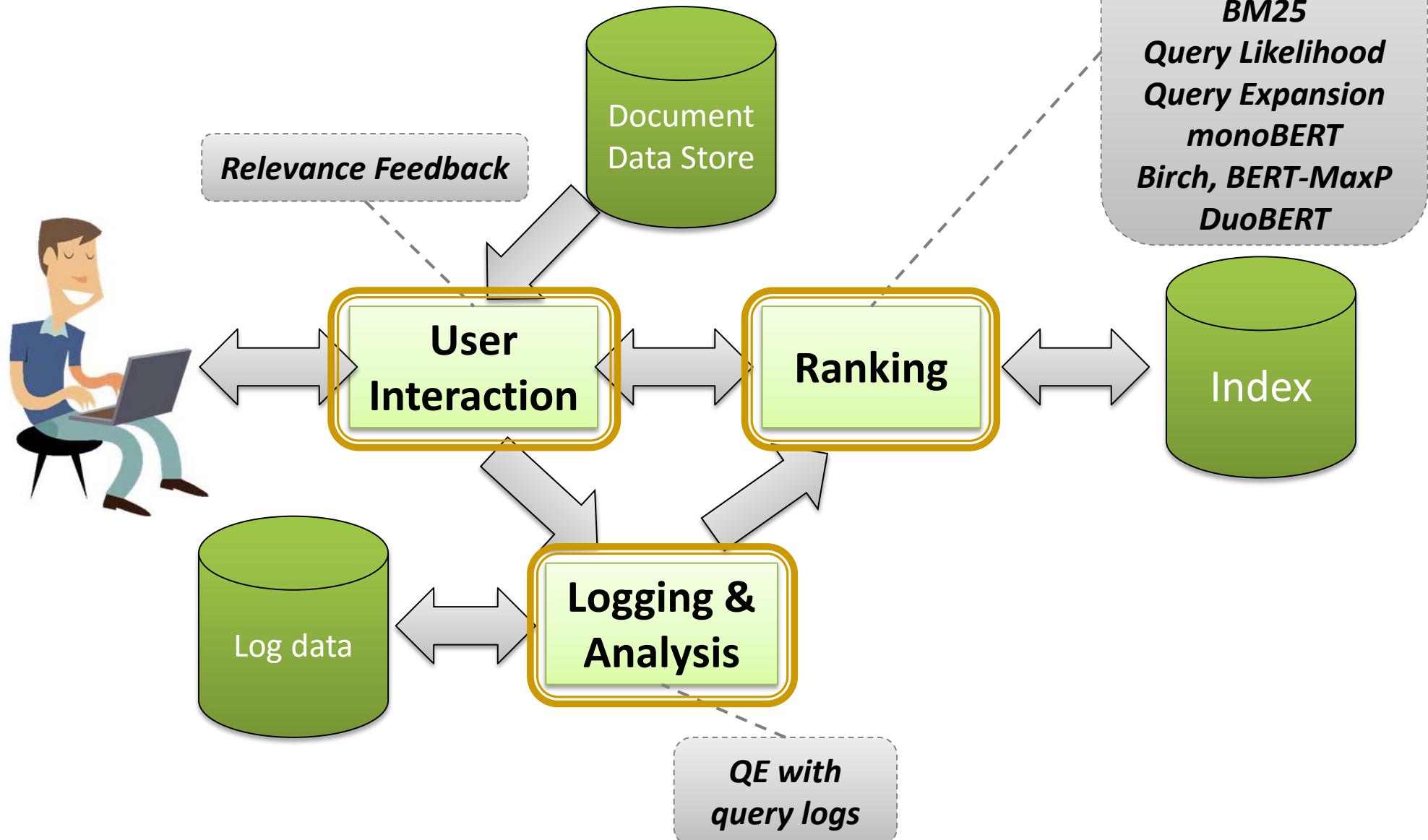
Inside the IR Black Box



Indexing process (offline)



Search process (online)

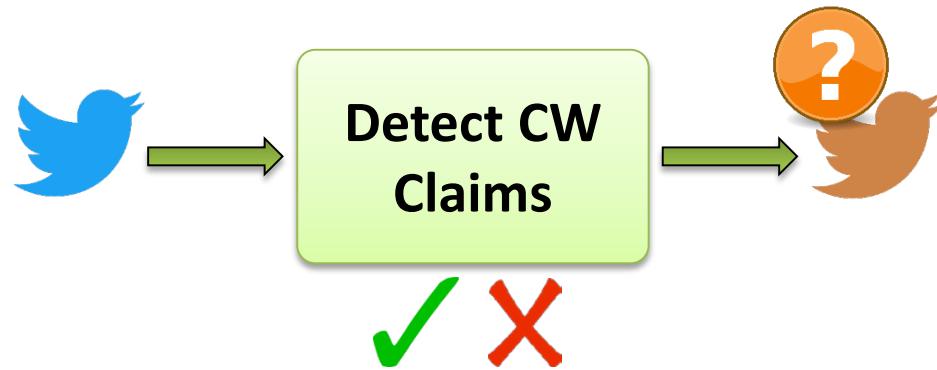


ONGOING IR RESEARCH ...



Tweet Check-worthiness

Problem: Given a stream of tweets (related to a topic), identify check-worthy (CW) claims.



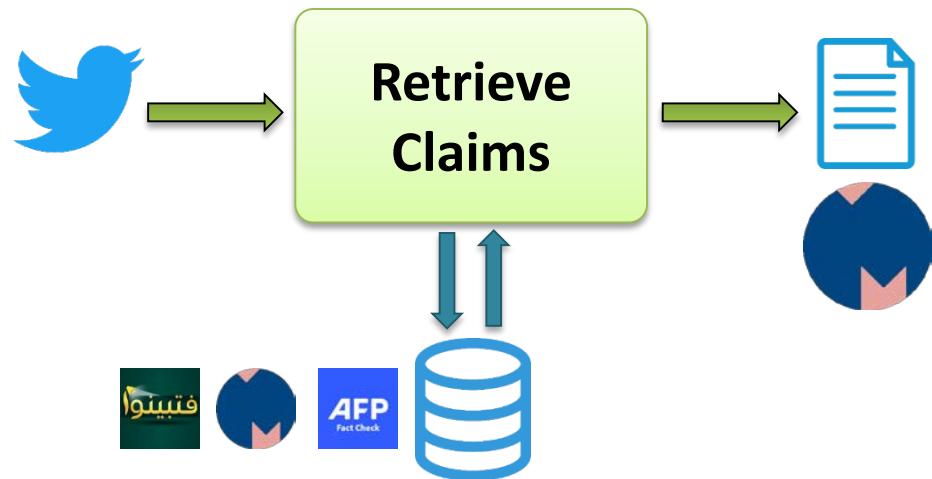
bigIR at CheckThat! 2020: Multilingual BERT
for Ranking Arabic Tweets by Check-worthiness

Maram Hasanain and Tamer Elsayed

Computer Science and Engineering Department, Qatar University, Doha, Qatar
{maram.hasanain, telsayed}@qu.edu.qa

Claim Retrieval

Problem: Given an [Arabic] tweet, retrieve claims already-verified by existing fact-checking organizations.

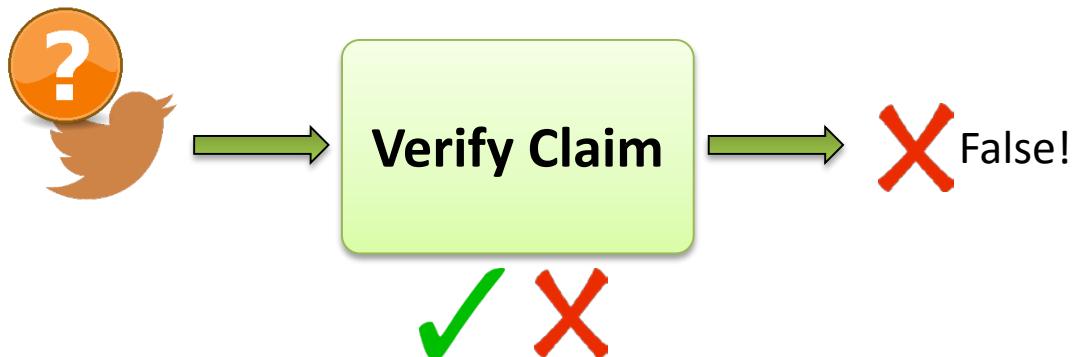


AraFacts: The First Large Arabic Dataset of Naturally-Occurring Professionally-Verified Claims

Zien Sheikh Ali, Watheq Mansour, Tamer Elsayed, and Abdulaziz Al-Ali
Computer Science and Engineering Department, Qatar University
{zs1407404, wm1900793, telsayed, a.alali}@qu.edu.qa

Claim Verification

Problem: Given a tweet containing a check-worthy claim, verify the claim.



ArCOV19-Rumors: Arabic COVID-19 Twitter Dataset for Misinformation Detection

Fatima Haouari, Maram Hasanain, Reem Suwaileh, Tamer Elsayed

Computer Science and Engineering Department, Qatar University

{200159617, maram.hasanain, rs081123, telsayed}@qu.edu.qa

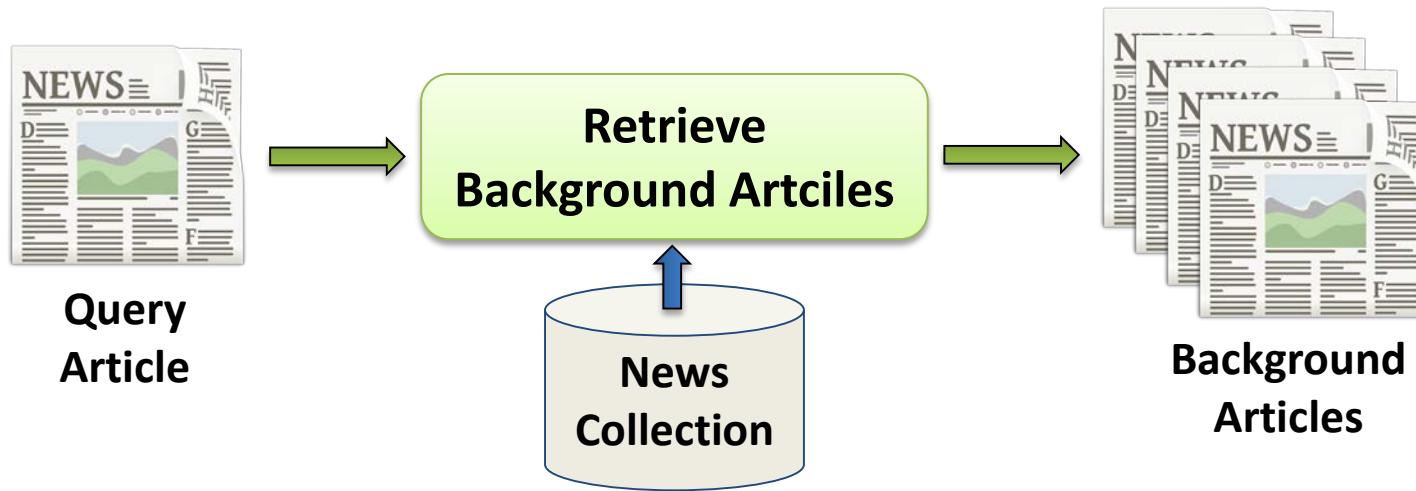
User Credibility

Problem: Given a twitter user, check if h/she is prone to spread fake news or not.



News Background Linking

Problem: Given a news article, find background articles that can help news-readers conceptualize the story of that given article with more knowledge.



Why is That a Background Article? A Qualitative Analysis of Relevance for News Background Linking

Marwa Essam
me1709534@qu.edu.qa
Computer Science and Engineering Department
Qatar University
Doha, Qatar

Tamer Elsayed
telsayed@qu.edu.qa
Computer Science and Engineering Department
Qatar University
Doha, Qatar

Time-Critical Geolocation for Social Good

- Given a tweet during crisis time, recognize location mentions and resolve them to locations on the map.



Are We Ready for this Disaster? Towards Location Mention Recognition from Crisis Tweets

Reem Suwaileh*

Muhammad Imran†

Tamer Elsayed*

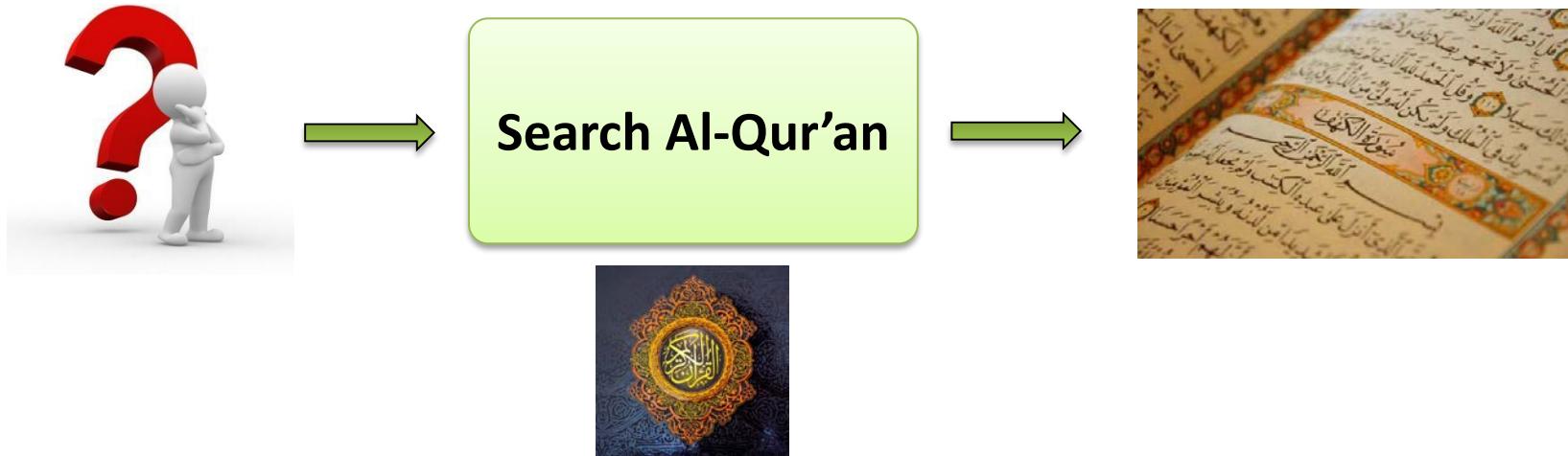
Hassan Sajjad†

*Computer Science and Engineering Department, Qatar University
{rs081123, telsayed}@qu.edu.qa

†Qatar Computing Research Institute, Hamad Bin Khalifa University
{mimran, hsajjad}@hbku.edu.qa

Qur'an QA

- Given an MSA question, find the answer in Al-Qur'an in words/verses.



AyaTEC: Building a Reusable Verse-Based Test Collection for Arabic Question Answering on the Holy Qur'an

RANA MALHAS and TAMER ELSAYED, Qatar University





OPEN IR RESEARCH PROBLEMS

TREC

<https://trec.nist.gov/>



TREC 2021

Clinical Trials Track

- focuses research on the clinical trials matching problem: *given a free text summary of a patient health record, find suitable clinical trials* for that patient.

Conversational Assistance Track

- forum for building and testing systems that *engage in open-domain information centric conversational dialogues*.

TREC 2021

Deep Learning Track

- focuses on IR tasks where a large training set is available, allowing us to compare a variety of *retrieval approaches including deep neural networks* and strong non-neural approaches, to see what works best in a large-data regime.

Fair Ranking Track

- focuses on fairly prioritizing Wikimedia articles for editing to provide a fair exposure to articles from different groups.

TREC 2021

Health Misinformation Track

- fosters research on retrieval methods that promote better decision making with *search* engines, and *promote reliable and correct information over misinformation.*

Incident Streams Track

- automatically processes social media streams during emergency situations with the aim of *categorizing information and aid requests* made on social media *for emergency service operators.*

TREC 2021

News Track

- develops test collections that *support the search needs of news readers and news writers* in the current news environment.

Podcasts Track

- develops methods for information *retrieval* and content understanding from open-domain podcast transcripts and audio.

RELEVANT JOBS



Giant Software Companies *with Search Services ...*

Google

Microsoft

Apple

Amazon

Twitter

Facebook

LinkedIn

IBM

booking.com, Airbnb, hotels.com, etc.

eBay

...

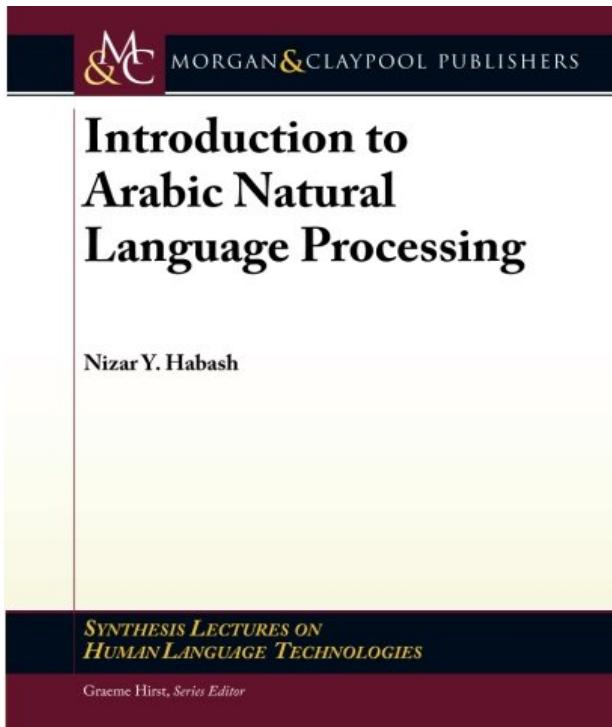


RESOURCES FOR ARABIC IR



Arabic IR



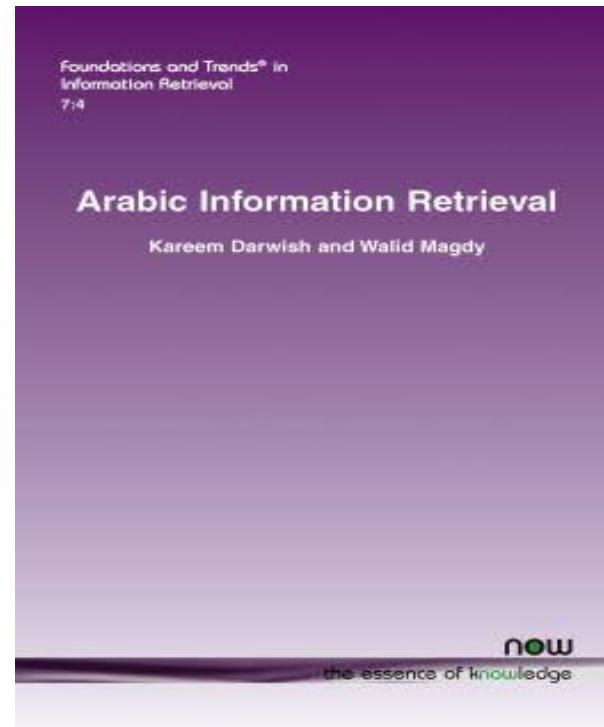


Introduction to Arabic natural language processing

[NY Habash](#) - Synthesis Lectures on Human Language ..., 2010 - morganclaypool.com

This book provides system developers and researchers in natural language processing and computational linguistics with the necessary background information for working with the Arabic language. The goal is to introduce Arabic linguistic phenomena and review the state ...

☆ 50 Cited by 750 Related articles All 17 versions



[HTML] Arabic information retrieval

[K Darwish, W Magdy](#) - Foundations and Trends in Information Retrieval, 2014 - dl.acm.org

In the past several years, Arabic Information Retrieval (IR) has garnered significant attention. The main research interests have focused on retrieval of formal language, mostly in the news domain, with ad-hoc retrieval, OCR document retrieval, and cross-language retrieval ...

☆ 9 Cited by 95 Related articles All 7 versions

Papers on Arabic IR at SIGIR

Search Results

Search Advanced Search

Applied Filters RSS

SIGIR X

Clear All

RESULTS VIDEOS Showing 1 - 19 of 19 Results

19 Results for: Publication Title: arabic 19 Edit Search Save Search

Searched The ACM Full-Text Collection (615,097 records) | Expand your search to The ACM Guide to Computing Literature (2,921,867 records)

Papers on Arabic IR at SIGIR

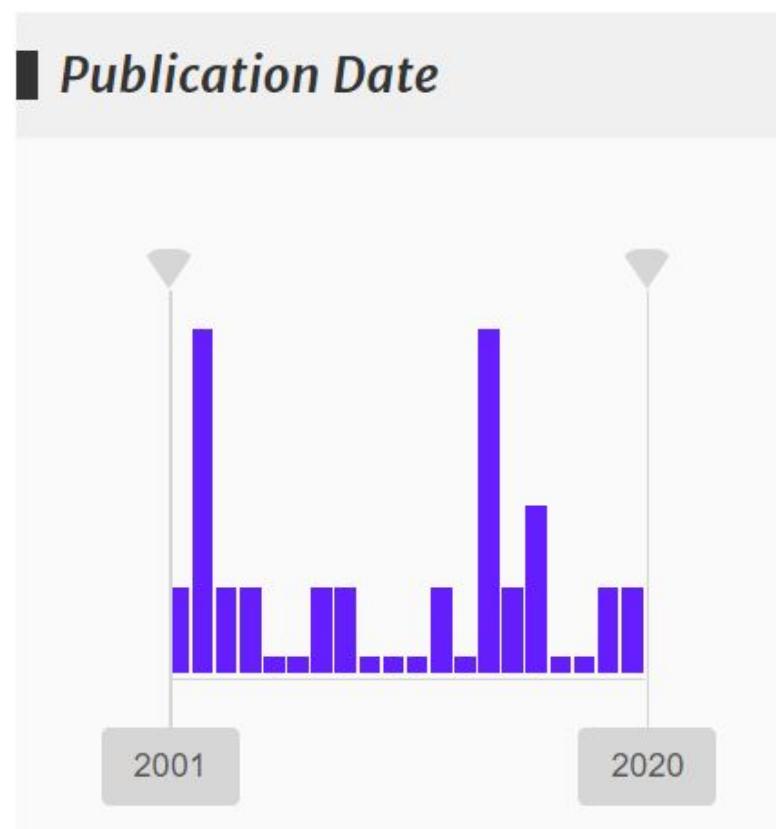
Conferences

Sponsors ▾

Conference Event ▾

Proceedings Series ▾

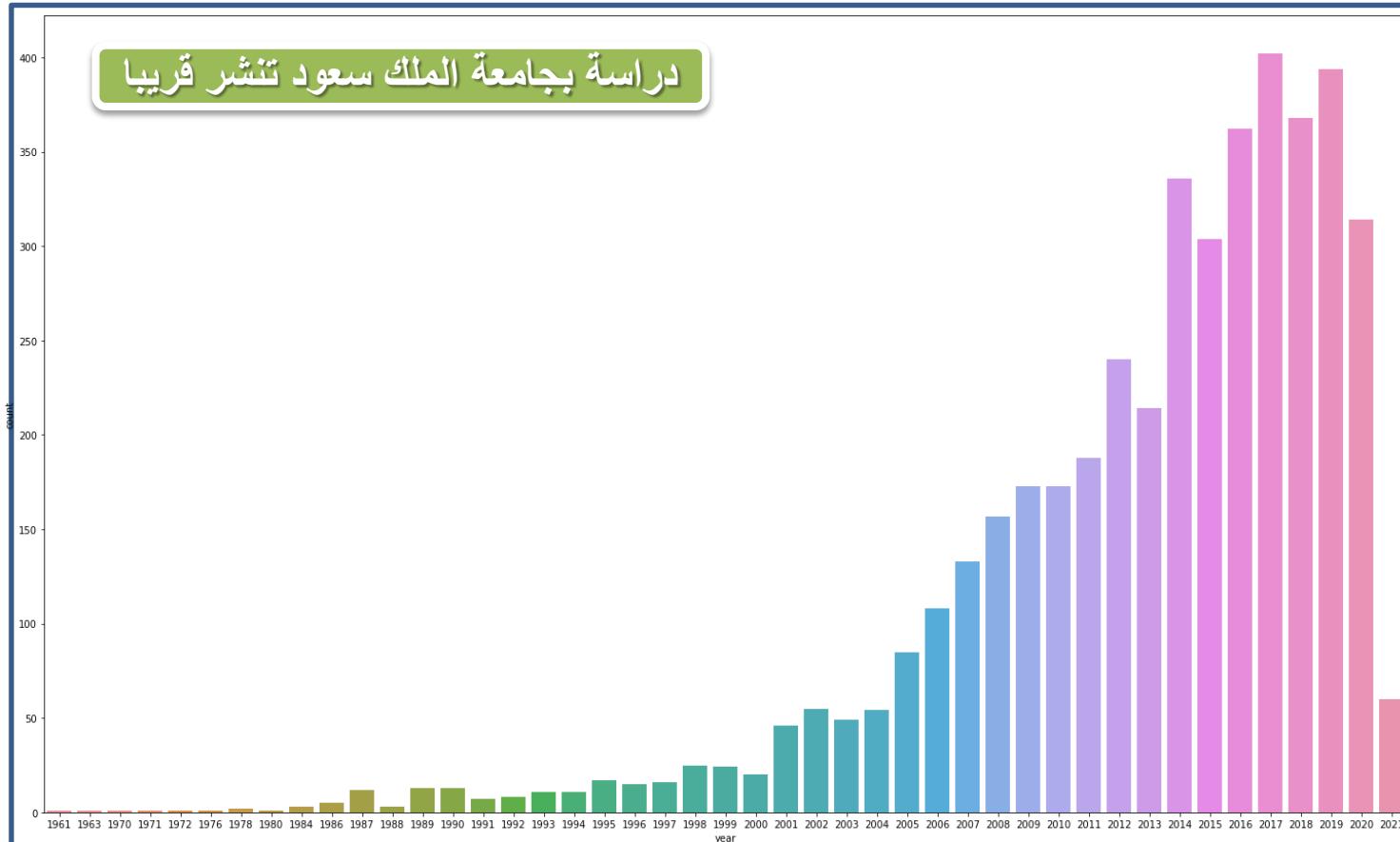
- IR: Research and Development in Information Retrieval (9)
- CIKM: Conference on Information and Knowledge Management (8)
- JCDL: Joint Conference on Digital Libraries (2)



Institutions ▾

- Qatar University (4)
- IBM Thomas J. Watson Research Center (2)
- Illinois Institute of Technology (2)
- Qatar Computing Research Institute (2)
- University of Maryland (2)
- University of Massachusetts Amherst (2)
- BBN Technologies (1)

Papers on Arabic NLP



TDT-2004: ADAPTIVE TOPIC TRACKING AT MARYLAND

Tamer Elsayed, Douglas W. Oard, David Doermann

Institute for Advanced Computer Studies

University of Maryland, College Park, MD 20742

Contact author: telsayed@cs.umd.edu

Gary Kuhn

National Security Agency

9800 Savage Road, Suite 6514

Fort Meade, MD 20755-6514

This is my first IR course in Arabic!

Challenges

- Not studied in most of our Arab universities!
- Online courses (YouTube, Coursera) are relatively few
- Many Arab researchers in this field do not work on the Arabic language!
- The datasets available for research and evaluation are very limited!







Arabic Web Search

1

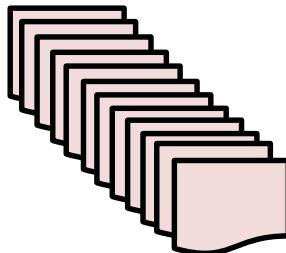
ArTest



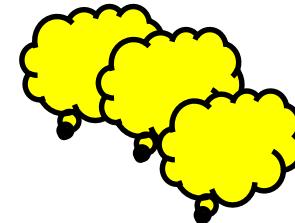
ArTest: The First Test Collection for Arabic Web Search with Relevance Rationales

Maram Hasanain^{*}, Yassmine Barkallah[†], Reem Suwaileh^{*}, Mucahid Kutlu[‡], Tamer Elsayed^{*}

^{*}Qatar University, Qatar [†]Innovation Lab, Mantu, Tunisia [‡]TOBB University of Economics and Technology, Turkey
 {maram.hasanain,rs081123,telsayed}@qu.edu.qa, yassmine.barkallah@outlook.com, m.kutlu@etu.edu.tr



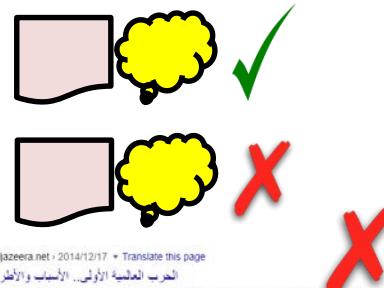
150M



نتائج الحرب العالمية الأولى



50



<https://www.aljazeera.net> • 2014/12/17 • Translate this page
الحرب العالمية الأولى، الأسباب والأطراف والمحاربون
كان الناس يشاركون العرب العالمية الأولى مملكة بريطانيا العظمى وبلجيكا ضد...
الروس في معركة بودا بودا في 1914...
Dec 17, 2014...
نحو 100,000 من الجنود العرب...
...

https://ar.wikipedia.org/wiki/%D9%82%D9%86_%D8%A7%D9%84%D8%A7%D9%85%D9%8A%D9%87%D9%8A • ترجمة هذه الصفحة
الحرب العالمية الأولى، الأسباب والأطراف والمحاربون
يشكل المعرضون الجدول الأولي تأثير حركة إنجلترا عام 1918 على نتائج الحرب. بالإضافة إلى النتائج الاقتصادية والسياسية...
الأسباب الإقليمية، وراء الأطماع، معاشر مارس، حصار غالينا...

10,529



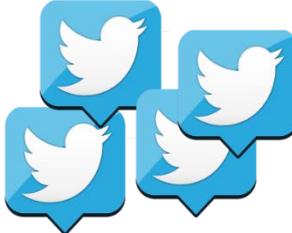
Arabic Twitter Search

EveTAR

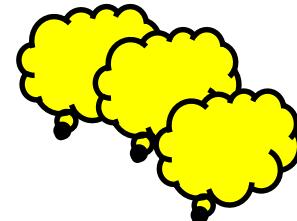


EveTAR: building a large-scale multi-task test collection over Arabic tweets

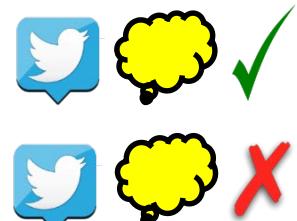
Maram Hasanain¹ · Reem Suwaileh¹ · Tamer Elsayed¹ ·
Mucahid Kutlu¹ · Hind Almerekhi²



355M



50



61,946



Qur'an Search

3

AyaTEC

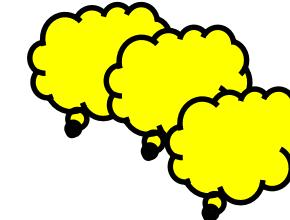


AyaTEC: Building a Reusable Verse-Based Test Collection for Arabic Question Answering on the Holy Qur'an

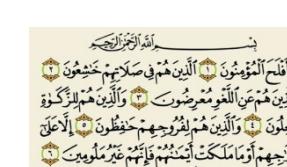
RANA MALHAS and TAMER ELSAYED, Qatar University



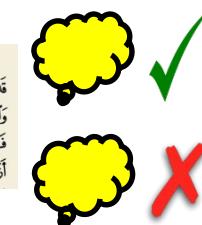
6200



207



1,762



Arabic Claim Retrieval

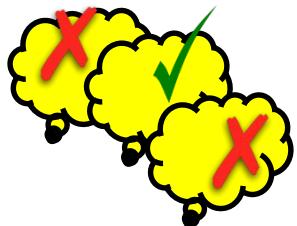
4

CheckThat! Claim Retrieval



The CLEF-2021 CheckThat! Lab
on Detecting Check-Worthy Claims,
Previously Fact-Checked Claims, and Fake News

Preslav Nakov¹, Giovanni da San Martino², Tamer Elsayed³,
Alberto Barrón-Cedeño⁴, Rubén Míguez⁶, Shaden Shaar¹, Firoj Alam¹, Fatima
Haouari³, Maram Hasanain³, Nikolay Babulkov⁵, Alex Nikolov⁵, Gautam
Kishore Shahi⁷, Julia Maria Struß⁸, Thomas Mandl⁹



30,329



599



649



Arabic COVID-19 Rumor Detection

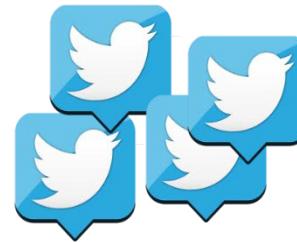
ArCOV19-Rumors

5

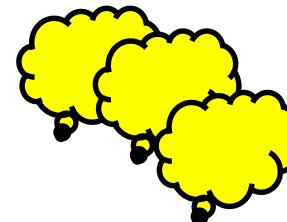


ArCOV19-Rumors: Arabic COVID-19 Twitter Dataset for Misinformation Detection

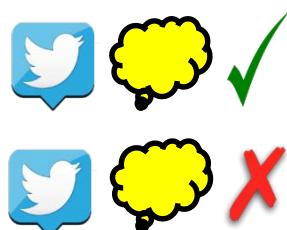
Fatima Haouari, Maram Hasanain, Reem Suwaileh, Tamer Elsayed
Computer Science and Engineering Department, Qatar University



2.7M



138



9,400



Best Paper Award

Awarded to

Fatima Haouari, Maram Hasanain,
Reem Suwaileh and Tamer Elsayed

for the paper

ArCOV19-Rumors: Arabic COVID-19 Twitter Dataset
for Misinformation Detection



Nizar Habash
WANLP 2021 General Chair

For more ...



<http://qufaculty.qu.edu.qa/telsayed/datasets/>

Google™

bing®

YAHOO!



FOR GRAD STUDENTS ...



تامر السيد

ما زال يعنى أن تكون
طالب دكتوراه ..

الدّوافع والجودة والمهارات الشخصية
والبيئة الداعمة

#ماذا_يعني_أن_تكون_طالب_دكتوراة



Tamer Elsayed
@tamer_elsayed

(1)
سأقوم إن شاء الله تعالى بالمشاركة بسلسلة تغريدات تحت عنوان

#ماذا_يعني_أن_تكون_طالب_دكتوراة

والهدف هو إبداء *رأي شخصي * في هذا الأمر لمساعدة الطلبة الحاليين والجدد وحتى الذين يفكرون في الالتحاق بهذا المسار ولتوسيع بعض الأمور الواجب إدراها مبكرا وفتح باب الحوار والنقاش فيها

[Translate Tweet](#)

10:14 AM · Mar 11, 2019 · Twitter Web Client

[View Tweet activity](#)

12 Retweets 55 Likes



Tamer Elsayed @tamer_elsayed · Mar 11, 2019

Replying to @tamer_elsayed

(2)
وسيكون التركيز بالأخص على طلبة الدكتوراة *في الجامعات العربية* وإن كان كثير من المفاهيم تنسحب على العموم أيضا.

أعذر إذا لم تكن مهتمة قد قررت أن تتوقف عن آخر ... وقد تم ... وأرجو أن يتم



PROJECT IDEAS



Some ideas for projects ...

- Dialectal Arabic web search
- Search engine for “Fatwas”
- Hadeeth verification over Twitter (social media)
- Arabic fact checking by finding evidence (on the Web)
- Arabic news classification
- Arabic employment recommendation/matching





WHERE CAN YOU GO FROM HERE?

Where can you go from here?

- Spread what you learned!
 - Teach it (all material is available)
- Read from the textbooks
 - Study other "basic" topics
 - Study advanced topics
- Read IR papers
 - Conferences: SIGIR, CIKM, Web, WSDM, ICTIR
 - Journals: TOIS, IP&M, IRJ, TKDE
- Do a project
- Participate in TREC (or a like)
 - CLEF has Arabic tasks

Reach us!



[@tamer_elsayed](https://twitter.com/tamer_elsayed)



[@Tamer Elsayed](https://www.youtube.com/c/TamerElsayed)



[telsayed AT qu.edu.qa](mailto:telsayed@qu.edu.qa)



<http://qufaculty.qu.edu.qa/telsayed/>



[@bigIR_group](https://twitter.com/bigIR_group)



<https://sites.google.com/view/bigir>



Acknowledgments

ALLAH

Fatima Haouari

Watheq Mansour

Maram Hasanain

My Family!



A graphic featuring the word "Feedback" written in white chalk-like font on a dark blue, textured background. The letter "F" has a yellow arrow pointing towards it from the top left, and the letter "k" has a yellow arrow pointing away from it towards the bottom right.

Feedback

