

# **Arabic News Summarization**

Julia Freeman and Souleiman Ayoub

# Overview

Customizing a filled out template provided by Fusion.

Summarize news articles with attributes that are extracted

- Named entities especially for organizations and persons
- Topic extraction using LDA algorithms
- Predefined news articles categories
- Dates

The data set are news articles extracted from PDF newspapers.

# Fusion

Fusion extends any Solr deployment with the enterprise-grade capabilities.

Shipped with many available OOB tools such as connectors.

Fusion is a full information retrieval system and supports building collections by

- Crawling (collecting) the data
- Indexing
- Searching



Collections



Search



Pipelines



Scheduler









Access



System

## Actions

Pipeline for: 
 Datasource: 
 Query Profile 
 Index Profile 

Query Pipelines 
 New Datasource 

## Status

Last modified 14-Apr-2015 12:59	Datasources	Documents 13604	Index size 913.87 MB	<input type="text" value="Search collection"/>
<b>Hard Commit</b>				
signals <input checked="" type="checkbox"/> ON	searchLogs <input checked="" type="checkbox"/> ON	dynamicSchema <input type="checkbox"/> OFF		

Clear collection

## Delete collection

- Delete Solr collection?
- Delete associated system collections?

Delete collection

# Schema

We modified the schema file by removing the unnecessary fields and adding an extra field.

This supports our special collection.

# Template Summarization

Fusion provides a template where each field contains an arbitrary descriptor.

Title

Author Name

Organization

Publication

date

People

Category

Topics

# Interface

*Arabic News Articles  
Template Summary*

Eg : Qatar, International, Graduate,



قطر

Al Arab Newspaper

Search

Eg : Qatar, International, Graduate,

Results 1 - 10 of 52245

Next 10 →

#### Article

لندن - قنا : قام سمو الشيخ تميم بن حمد آل ثاني ولي العهد رئيس اللجنة الاولمبية القطرية امس بزيارة امس

[See More](#)

#### Summary

العنوان: تميم بن حمد آل ثاني ولي العهد رئيس اللجنة الاولمبية القطرية امس بزيارته الى بيت

تاريخ النشر: ٩/اغسطس/٢٠١٢

الكاتب: الشيخ تميم بن حمد آل ثاني

الأشخاص المشار إليهم: لشيخ تميم بن حمد آل ثاني

المؤسسات المشار إليها: اللجنة الاولمبية القطرية, قطر الوطني, قطر للعلوم, قطر الوطني, اتصالات قطر كيوتل, القطرية للاستثمار العقاري, متاحف قطر ومعهد الدوحة, العامة للسياحة

التصنيف العام: الرياضة

الكلمات في الموضوع الرئيسي: قطر, لندن, القطرية, سمو, اللجنة, الاولمبية, الرياضي, اولمبياد, دولة, تحقيق

[See Less](#)

#### Article

الدوحة - [ : اعلنت دار بلومزبري - مؤسسه قطر للنشر- عن تشييد سلسله «اكتشف بينه قطر

[See More](#)

#### Summary

«العنوان: حشرات وعناكب قطر

تاريخ النشر: ١٢/نوفمبر/٢٠١٣

الكاتب: اسم الكاتب غير متوفر

الأشخاص المشار إليهم: غير متوفر

المؤسسات المشار إليها: وسسه قطر للنشر, اوكسيدنتال قطر, اوكسي قطر, بمركز الدوحة

التصنيف العام: الفن و الثقافة

الكلمات في الموضوع الرئيسي: قطر, سلسله, بينه, اجزاء, بلومزبري, اكتشف, دار, الدوحة, مؤسسه, نشر

[See Less](#)

#### Article

صناعات قطر - بنك خليجي - مصرف الريان - الوطني للاجازة - قطر للتأمين - شركة الكهرباء والماء القطرية -

[See More](#)

#### Summary

العنوان: شركة الخليج القابضة -الخليج الدولي للخدمات - صناعات قطر - بنك خليجي - مصرف الريان

تاريخ النشر: ٦/مايو/٢٠١٠

الكاتب: اسم الكاتب غير متوفر

الأشخاص المشار إليهم: غير متوفر

المؤسسات المشار إليها: لقطرية البنك التجاري بنك الدوحة شركة الخليج, الخليج الدولي للخدمات صناعات قطر بنك, مصرف الريان الوطني, قطر للتأمين شركة, والماء القطرية,

مصرف قطر الإسلامي بنك قطر الدولي الإسلامي, القطرية كيوتل, الدولي المتحد للتعمية

التصنيف العام: الاقتصاد

الكلمات في الموضوع الرئيسي: شركة, قطر, معايير, الخليج, الدولي, مصرف, القطرية, الإسلامي, الاختيار, خمسة

[See Less](#)



# Acknowledgements

Professor: Dr. Edward Fox

Client: Tarek Kanan

QNRF – NPRP Grant #: 4-029-1-007