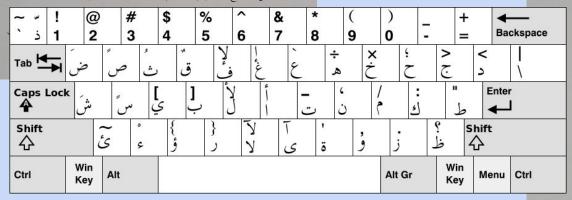


Ranking Arabic Documents with Machine Learning



- [14] أسر أبي الفرج وذلك لعدم احتياطه عندما تأخر عليه صاحبه ولم يأتي كان الأولى في هذه الحالة أن يرسل من ينوبه .
- (15) أسر شريف الله المصري بسبب تعامله مع أناس هو يعلم أنحم مخترقين ولهم صلة بال isi .
- (16) مقتل عكرمة وإخوانه بسبب عدم تغيير البيت بعد أن تم كشفه وقد غدا معلوماً للقاصي والداني .
- {17} مقتل الشيخ الكندي وذلك لعدم أخذه بجدية ما جاءه من نية الباكستان الهجوم



- 28 Letter Alphabet
- Logical structure (nouns and verbs) similar to English
- Written and read from right to left.
- Written in cursive, making text recognition challenging.

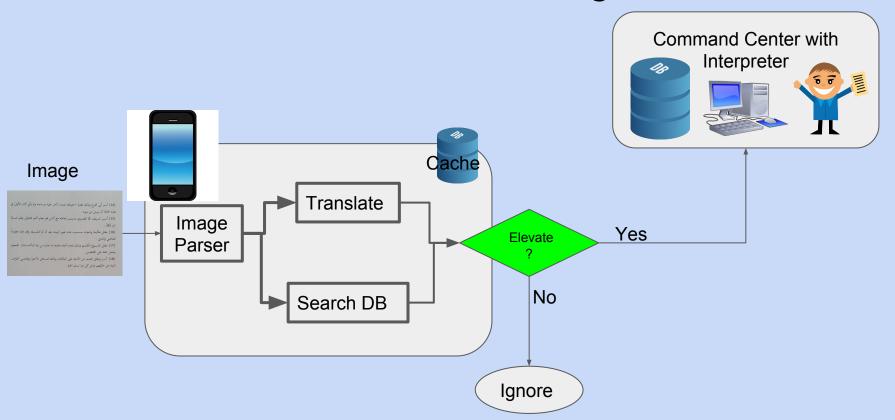
- وَيُؤَوُّهُمْ حَرْبُ الشَّيْمَةُ ؟ - ولماذا لا يُعاسَبُ الطَّامُ العلويُّ على جرائمه الكثيرة في لبنانُ وسوريَّةُ؟
- جَمْهُ إِنْ الْمِيمَ الْمُنْهُ وَمُونَا اللَّهِ اللَّهِ اللَّهِ اللَّهِ عَلَيْهِ عَلَيْهِ عَلَيْهِ المُنافِعِ ا - ولعاذا لا تُرَدُّ خَوْلُ الْأَمْهِاتِ اللَّكِ اللَّالِي اللَّالِي اللَّهِ عَلَيْهِ عَلَيْهِ عَلَيْهِ اللَّهِ عَلَيْهِ عَلَيْهِ اللَّهِ عَلَيْهِ عَلَيْهِ عَلَيْهِ اللَّهِ عَلَيْهِ عَلَيْهِ عَلَيْهِ عَلَيْهِ اللَّهِ عَلَيْهِ عَلَيْهِ عَلَيْهِ عَلَيْهِ عَلَيْهِ عَلَيْهِ عَلَيْ
- في الطُرْقاتِ وَوَكُوا تَجْمَقْنَهُم دَمَاؤَهُم؛ على يَدِ هذه الطَّانِهُ عَلَيْهِمُ الطَّانِةِ وَاذَرْكِمْ فِي - ولعاذا لا يُعاسَبُ مِن قبل أهل السنةِ في السابع من آيار، مع أنَّ يعضُمُ الآفِيمَ بِي كَاللَّهُ - ولعاذا لا يُعاسَبُ مَن قبلُ أهل السنةِ في السابع من آيار، مع أنَّ بعضُمُ الآفِمِ المُنتَجِعُ عَلَيْهُمُ المُنتَانِعُ عَلَيْهُمُ المُنتَانِعُ عَلَيْهُمُ المُنتَانِعُ عَلَيْهُمُ المُنتَانِعُ عَلَيْهُمُ المُنتَانِعُ عَلَيْهُمُ المُنتَانِعِيمُ عَلَيْهُمُ المُنتَانِعُ عَلَيْهُمُ المُنتَانِعُ عَلَيْهُمُ عَلَيْهُمُ عَلَيْهُمُ عَلَيْهُمُ عَلَيْهُمُ عَلَيْهُمُ عَلَيْهُمُ السِنْعُ عَلَيْهُمُ عَلَيْهُمُ عَلَيْهُمُ عَلَيْهُمُ عَلَيْهُ عَلَيْهُمُ عِلَيْهُمُ عَلَيْهُمُ عَلَيْ
- ولعاذا لا يُضَح تحقيقُ في مصير مثاتِ ممن قُبلوا مِن التعليبِ في السجون، ومَن قُبلوا بمم باردٍ
- اماذا يتجاهلون ما وقع عليكم يا أهل السنة من قبيح الظلم فعظيم الجرائم؟
- لماذا لا يستجاب لدعاوى رفع الظلم إلا إذا كان هذه الدعاوى من يجم الطائفة الشيعية؟ : تِ كِالسَّمَةِ وَلِمِلمَّا إِلَمُهُ إِنَّ فِينَسَا الْمُهُ لِيلِمُهُ لَا لِمُحْلًا يَوْمُوا لَنْ وَقِيهِ فَعَلَمُ اللَّهِ عَلَيْهُ لِمَا السَّبَّةِ وَلِمُعُوا تَمِينُهُ لِمَا اللَّهُ اللَّهِ اللَّهُ عَلَيْهُ لِمَا اللَّهُ عَلَيْهُ لَمَا اللَّهُ عَلَيْهُ لَمَا اللَّهُ عَلَيْهُ لِمَا اللَّهُ عَلَيْهُ لِمَا اللَّهُ عَلَيْهُ لَمَا اللَّهُ عَلَيْهُ لَمَا اللَّهُ عَلَيْهُ لَمَا اللَّهُ عَلَيْهُ لَلَّهُ اللَّهُ لَلَّهُ عَلَيْهُ لَلَّهُ اللَّهُ لَلَّهُ لَلَّهُ لَلَّهُ لَللَّهُ لَلْمُ لَا لَهُ عَلَيْهُ لَلَّهُ لَلْمُ لَا اللَّهُ لِمُعْلِقًا لَمْ عَلَيْهُ لَللَّهُ لَلَّهُ لَلْمُ لَا لَهُ لَا لَهُ لَا لَهُ لَا لَهُ لَاللَّهُ لِمُعْلِقًا لِمُعْلِقًا لِمُعْلِقًا لِمُعْلِقًا لِمُعْلَقًا لِمُعْلَمُ لِمُعْلِقًا لِمُعْلِقًا لِمُعْلِقًا لِمُعْلِقًا لِمِنْ لَمُ لَهُ لَلَّهُ لَمُ لَلَّهُ لَلْمُ لَمِنْ لِمُؤْلِقًا لِمِلْ لَا لَهُ لَلَّهُ لَلْمُ لَا لَهُ لَا لَهُ لَلَّهُ لَلَّهُ لَا لَهُ لَا لَمُؤْلِقًا لِمُعْلَمُ لَلَّهُ لِمُعْلِقًا لِمُعْلِمُ لِمُعْلِقًا لِمُعْلِقًا لِمُعْلِقًا لِمُعْلِقًا لِمُعْلِقًا لِمِنْ لِمُعْلِقًا لِمِنْ لِمُعْلِقًا لِمُعْلِقًا لِمُعْلِقًا لِمُعْلِقًا لِمُعْلِقًا لِمُعْلِقًا لِمُعْلِقًا لِمِنْ لِمُعْلِقًا لِمُعْلِقًا لِمُعْلِقًا لِمُعْلِقًا لِمُعْلِقًا لِمُعْلِقًا لِمُعْلِقًا لِمُعْلِقًا لِمِنْ لِمِنْ لِمُعْلِقًا لِمُعْلِقًا لِمُعْلِقًا لِمُعْلِقًا لِمِنْ لِمُعْلِقًا لِمُعْلِقًا لِمُعْلِقًا لِمُعِلَّا لِمُعْلِقًا لمُعْلِقًا لِمُعْلِقًا لِمُعْلِقًا لِمُعْلِقًا لِمُعْلِقًا لِم وظلم في الدنيا؛ على يَدِ الطوائقِ المهيمنة على عبادِ اللهِ في أرضِ الشام، الناهبةِ لقراهِما، المفسدةِ لها؛ من للظاومة المستضعفة في هذا الزمان، فالظلم ثازل مجم بكل صوره، وبأعظم مستويات؛ فهو ظلم في الدين الأفراد .ونحن إذا نظرًا إلى أحوالِ أهلنا في لبنان وسوريَّة وسائر بلادِ الشام؛ وَحدْنًا أتُّهم مِن أعظم الطوائف الظالم ورمَّه أيُّ كان وسواءً كان الظالم نظامًا أم مؤسسة حكومية أم حولًا لم ظائمة أم كان فردًا مِن مسلما أو كافرا، ولا بكون الظلوم مسلماً أو كافراً؛ فيحبُ نصرة الظلوم أيًّا كان دينه وجنسُه، وبجبُ دفع في قوله تعالى (يبكم) يوحع إلى قوله (عبادي) فهو عامٌّ يشمل كلُّ عباد الله؛ فلا عبرة هنا بكون الظامِّ قال: (يا عبادي، إني حوَّثُ الطَّلَمُ على نفسي، وجعلتُه بينكُم محرِّمًا؛ فلا نظللوا ...) الحديث .والضعير مسلم عن أبي ذُرٍّ -رضي الله عنه- عن النبيُّ - صلى الله عليه وسلم- فيما يرويه عن ربَّه جل وعلا أنه فشريعة الله تحرُّم الطَّلمُ أبِين، لا تفرُّق في ذلك بين مؤمن وكافر، أو عربي وأعجميًا؛ وقد جاء في صحيح على أهل النمة؛ وحب الجهادُ لذلك، وبَلْدَك في سبيلِ ذلك نفوسُ المسلمينُ وأمواهُم؛ حتى لا تكونُ فتنه، وأموالهم وأعراضهم، ونصرة مظلومهم، ولانتصار لضعيفهم، حتى قال بعض أهل العلم: إنَّه إن اعتدى معتدر ظلم فيه لمؤمن ولا كافر، بل كأمهم يعيش مستوفيا حقوقه التي تكفلها له شريعة الله، في حفظ لغوسهم المسلمين، وإذا لم يكن على عبار الله – مؤمنهم وكافرهم- سلطانٌ غيرُ سُلطان الله العَمَال الحَقَّر؛ الذي لا ويكون الدين لله؛ إذا ارتفع تسلُّطُ الظامةِ – مِن الباطنية الشيعة وغيرهم- عن أهل الإسلام وبلاد

Conops

- Understand the value of the document
- Limited Interpreters upshoot documents that have urgent relevance
- Mobile app translation is nice to have



Architecture Design



User Interface

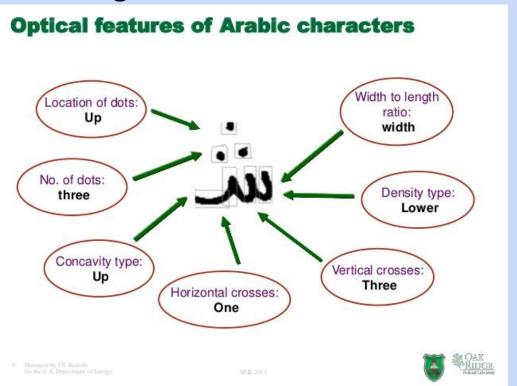
Demonstrate Image Parser



Arabic Handwriting Recognition
Using Neural Network in MATLAB

(Word Level)

Jonani Al-Inhibi



Problem Statement

- Given a document photo, what other documents is it like?
- Should someone translate it?

صدی 10/13 pt

من عداد الصنائع الإنسانية وهو رسوم وأشكال حرفية تدل على الكلمات المسموعة الدالة على ما في النفس. فهو ثاني رتبة عن الدلالة اللغوية وهو صناعة شريفة إذ الكتابة من خواص الإنسان التي يميز بها عن الحيوان. وأيضاً فهي تطلع على ما في الضمائر وتتأذى بها الأغراض إلى البلد البعيد فتقض الحاجات وقد دفعت مؤونة المباشرة لها ويطلع بها على العلوم والمعارف وصحف الأولين وما كتبوه في علومهم وأخبارهم فهي شريفة بجميع هذه الوجوه والمنافع... فالخط المجرد كماله أن تكون دلالته واضحة بإبانة حروفه المتواضعة وإجادة وضعها ورسمها كل واحد على حدة متميز عن الاخر إلا ما اصطلح عليه الكتاب من إيصال حرف

10/13 pt Seria Italic

Writing is the outlining and shaping of letters to indicate audible words which, in turn, indicate what is in the soul. It comes second after oral expression. It is a noble craft, since it is one of the special qualities of man by which he distinguishes himself from the animals. Furthermore, it reveals what is in (people's) minds. It enables the intention (of a person) to be carried to distant places.... It enables (people) to become acquainted with science, learning, with the books of the ancients, and with the sciences and information written down by them. Because of all these useful aspects, writing is a

Limitations and Assumption

Limitations:

- Small time and memory constraints (mobile computer)
- Limited connectivity to central database
- Limited central intelligence about documents in other language

Assume:

- Documents can be parsed into words
- Words can be mapped to their English word/words (Dictionary lookup)
- English words can be counted to create a histogram of words (BOW)



Demonstrate Classifier Solution

Key Idea:

- Convert document to a simple representation in native language
 - Represent documents as bag of words (BOW) histograms
- This allows train a classifier(s) using corpus of labeled documents in native language (HUGE CORPUS FROM NATIVE INTELLIGENCE)
- Classifiers quickly tell what kind of document it is like.
 - Optionally, also classify as "Important for further detailed translation"

Benefits:

- Simple classifiers on BOW histograms are just linear feature vectors
 - Low memory, ultra fast decisions
- As more intel is gained, the decision rule(s) can be updated

What Classifier?

- Classifier 1: Is it important for detailed translation by human?
- Classifier 2: What type of document is it like?/Which documents that I know are like it?
- You want:
 - Decision confidence
 - Fast lightweight decision
 - Small data decision update
 - <u>Nearest neighbor:</u> Distance implicitly tells you confidence about what type of document it is
 - Relevance vector machine (RVM): Gives you the category as well as explicit confidence score [0,1]
 - Simple weight vector (super small memory)

Limitations

Hidden assumption:

 BOW for word -> word translation will give proper representation for classifying using our English corpus.

If not:

- Learn an intermediate mapping that re-weights other language BOW to native language BOW.
 - Simple least squares reweighting of histogram
 - Cross correlational analysis

Summary

List of Resources

- http://arxiv.org/pdf/1206.1518.pdf research paper for recognizing Arabic characters
- http://www.amara.org/en/videos/cl4VRKdprrGv/info/arabic-handwriting-recognition-using-neuralnetwork-in-matlab-word-level-user-manual/ - video tutorial to use Matlab's Arabic Handwriting Recognition feature
- https://console.ng.bluemix.net/catalog/services/visual-recognition/ specs for Watson's Visual Recognition service

Backup Slides

ضاع في الترجمة

May 22, 2016

Safe Sign Group

Language Recognition

- There are thousands of languages spoken on Earth, but the top ten languages cover most speech.
- In field operations, English would be expected to be the default language in only a small percentage of situations.
- Arabic is spoken by

Modern Standard Arabic

 Challenges of translating content, and also obtaining the right context.

Simple Yet Complicated

