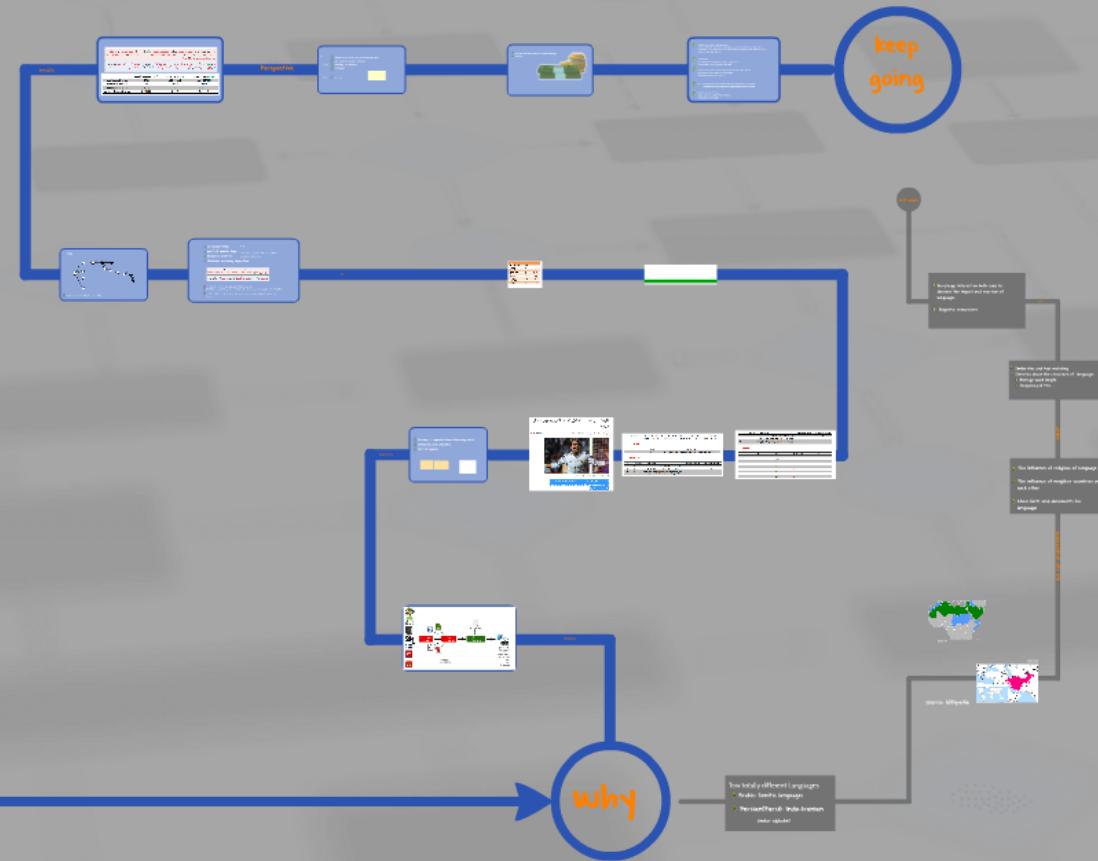


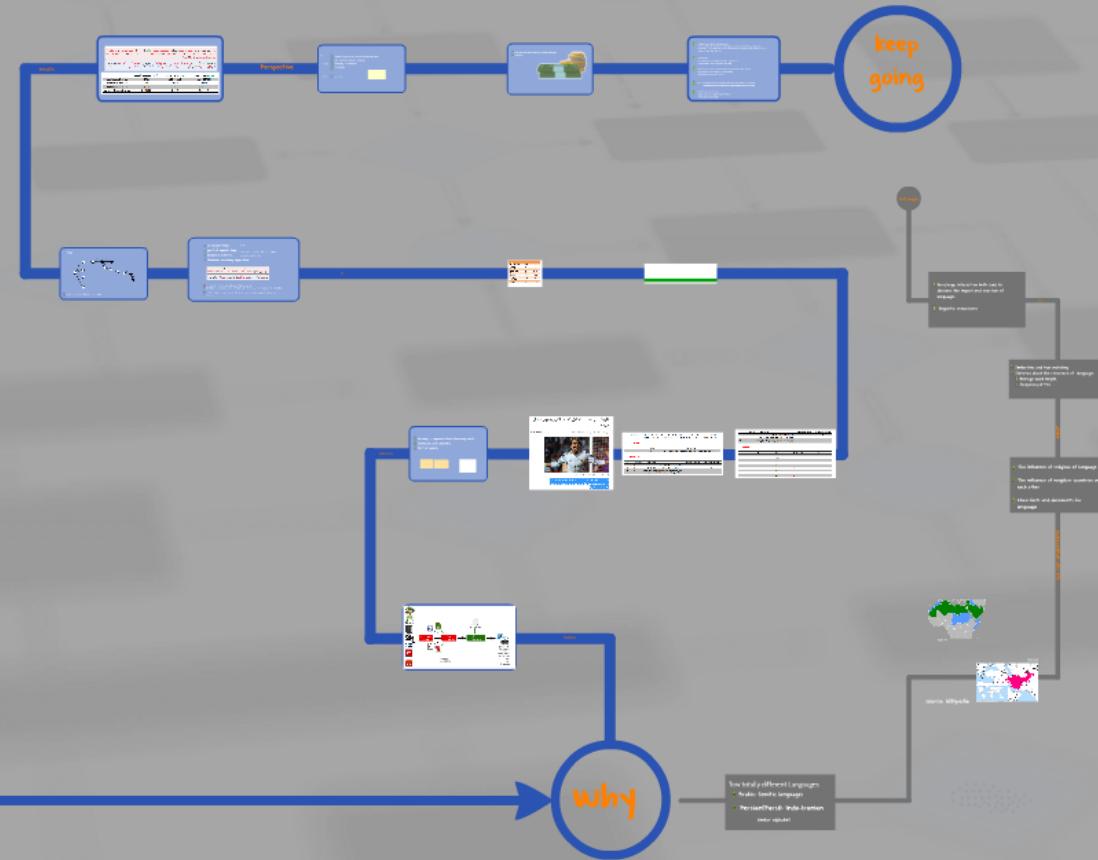
Analyzing and comparing
Arabic and Persian

Tariq and Iman



Analyzing and comparing
Arabic and Persian

Tariq and Iman



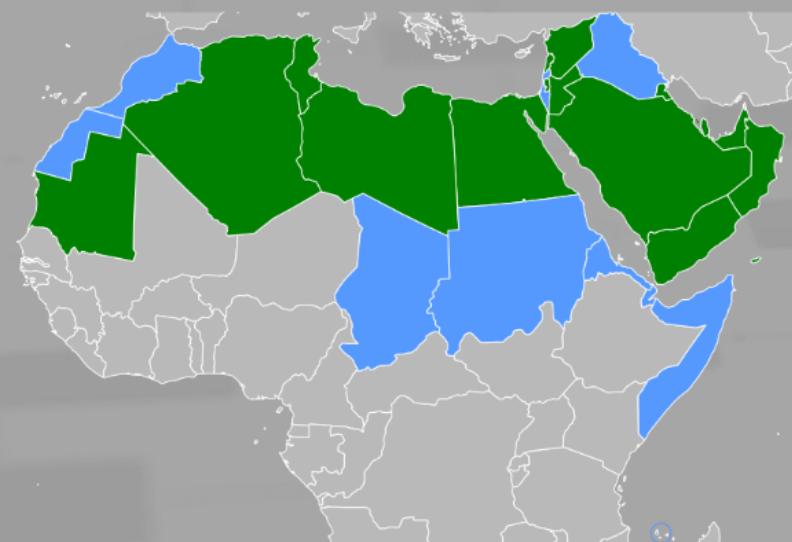
Analyzing and comparing
Arabic and Persian

Tariq and Iman

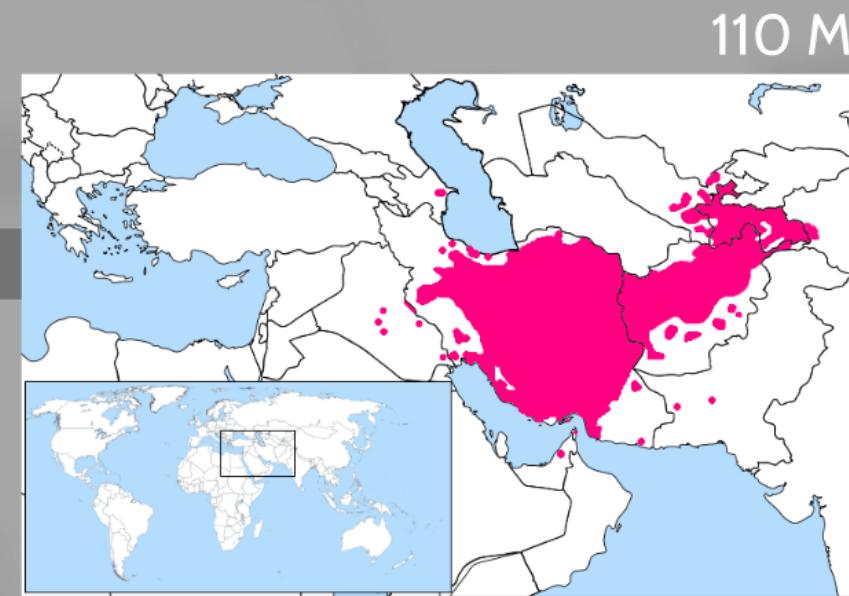
why

Tow totally different Languages

- 📍 Arabic: Semitic languages
- 📍 Persian(Farsi): Indo-Iranian
 - similar alphabet



420 M



110 M

we can understand

- || The influence of religious of language
- || The influence of neighbor countries on each other
- || More facts and documents for languages



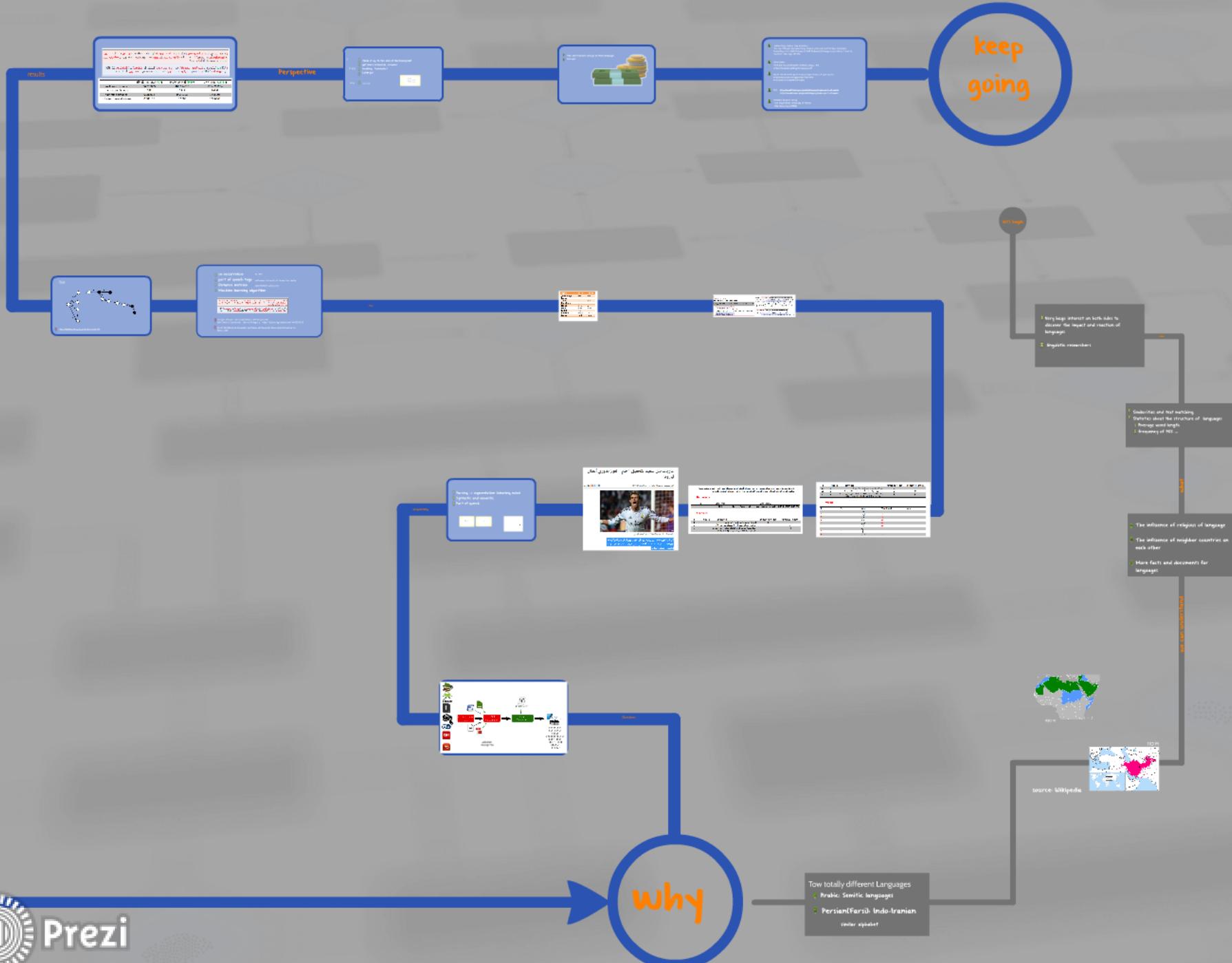
- | Similarities and text matching
- | Statistics about the structure of languages
 - | Average word length
 - | frequency of POS ...

WILHELM

- I Very huge interest on both sides to discover the impact and reaction of languages
- I linguistic researchers

let's begin

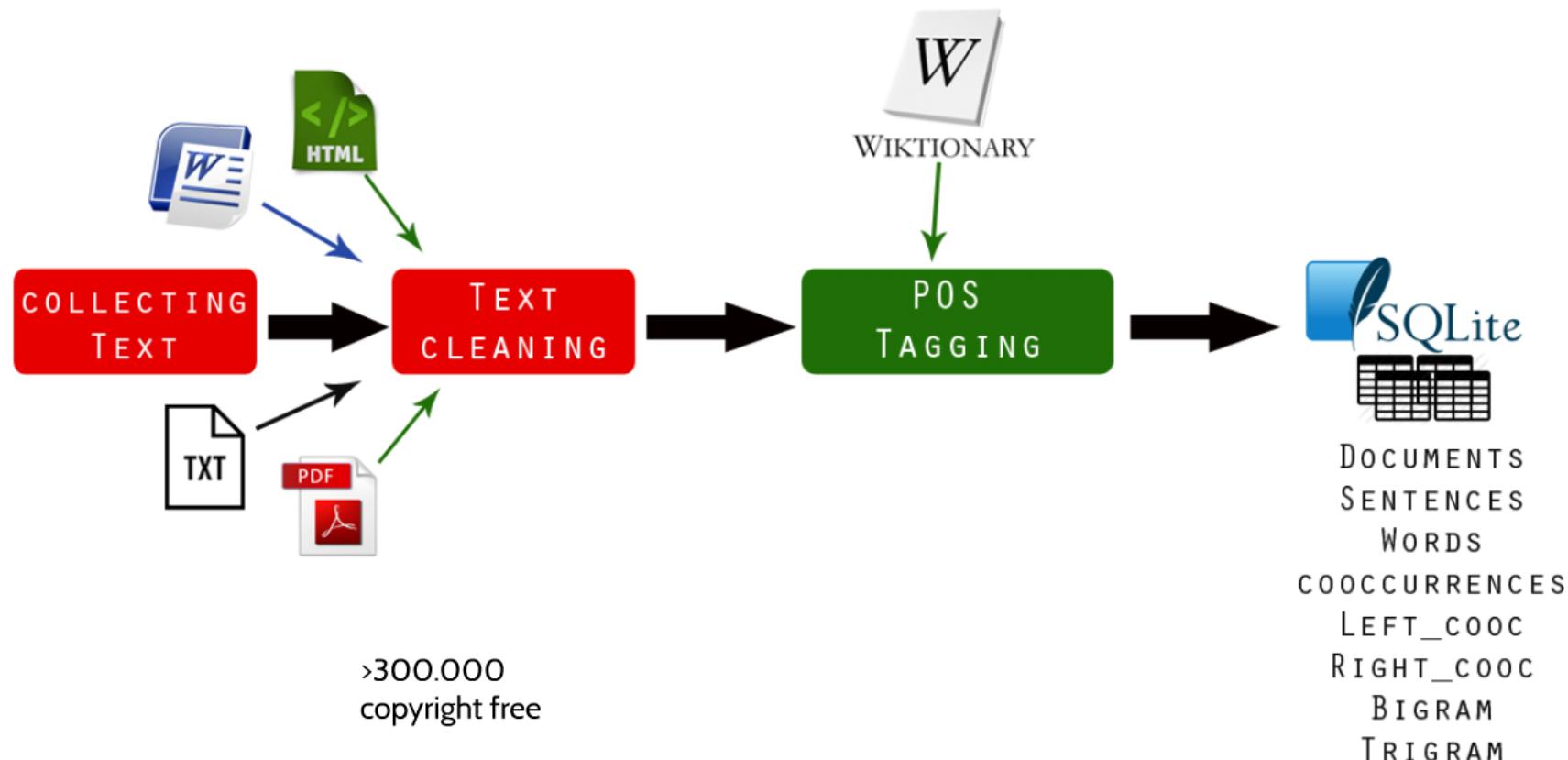




Database



Guran



prepossessing



Prezi

- | Parsing -> segmentation (cleaning noise)
- | Syntactic and semantic
- | Part of speech

/men/
/man/

من

كتاب

Germany

المانيا
المان

Germany

المانيا
المان

/men/
/man/ من

الكتاب
كاب

غاريث بيل سعيد بتحقيق "حلم" الفوز بدوري أبطال أوروبا

آخر تحديث: الأحد، 25 مايو / أيار 2014 GMT 13:31



الفعل بيل إلى ريال مدريد في صنفه قياسية الصيف الماضي

قال غاريث بيل، نجم فريق ريال مدريد، إن الفوز بدوري أبطال أوروبا سيظل معه "إلى الأبد". وذلك بعدما ساعد هدفة النادي الملكي على هزيمة أتليتيكو مدريد 1-0 في المباراة النهائية بالبطولة.

قال غاريث بيل، نجم فريق ريال مدريد، إن الفوز بدوري أبطال أوروبا سيظل معه "إلى الأبد". وذلك بعدما ساعد هدفة النادي الملكي على هزيمة أتليتيكو مدريد 1-0 في المباراة النهائية بالبطولة.

Documents:

id	doc_type	doc_name
1	URL	http://www.bbc.co.uk/arabic/sports/2014/05/140525

Sentences:

id	doc_id	sentence	sentence_order	sentence_length
1	1	قال غاريث بيل، نجم فريق ريال مدريد، إن الفوز بدوري أبطال أوروبا سيظل معه إلى الأبد.	1	7
2	1	إن الفوز بدوري أبطال أوروبا سيظل معه إلى الأبد.	2	9
3	1	وذلك بعدما ساعد هدفة النادي الملكي على هزيمة أتليتيكو مدريد 1-0 في المباراة النهائية بالبطولة.	3	14

id	doc_id	sentence	sentence_order	sentence_length
1	1	قال غاريث بيل، نجم فريق ريال مدريد،	1	7
2	1	إن الفوز بدوري أبطال أوروبا سيظل معه إلى الأبد.	2	9
3	1	وذلك بعدما ساعد هدفة النادي الملكي على هزيمة أتليتيكو مدريد 1-0 في المباراة النهائية بالبطولة.	3	14

Words:

Id	Sen_id	Word	Word_order	pos
1	1	قال	1	
2	1	غاريث	2	
3	1	تل	3	
4	1	نعم	4	
5	1	فريق	5	
6	1	ريال	6	
7	1	مدريد	7	
8	2	إن	1	
9	2	الفوز	2	

غاريث بيل سعيد بتحقيق "حلم" الفوز بدوري أبطال أوروبا



آخر تحديث: الأحد، 25 مايو / أيار، 2014، 13:31 GMT



Getty

انتقل بيل إلى ريال مدريد في صفقة قياسية الصيف الماضي

قال غاريث بيل، نجم فريق ريال مدريد، إن الفوز بدوري أبطال أوروبا سيظل معه "إلى الأبد"، وذلك بعدما ساعد هدفه النادي الملكي على هزيمة أتلتيكو مدريد 1-4 في المباراة النهائية بالبطولة.

قال غاريث بيل، نجم فريق ريال مدريد، إن الفوز بدوري أبطال أوروبا سيظل معه "إلى الأبد"، وذلك بعدما ساعده هدفه النادي الملكي على هزيمة أتلتيكو مدريد 1-4 في المباراة النهائية بالبطولة.

Documents:

id	doc_type	doc_name		
1	URL	http://www.bbc.co.uk/arabic/sports/2014/05/140525		
id	doc_id	sentence	sentence_order	sentence_length
1	1	قال غاريث بيل، نجم فريق ريال مدريد، إن الفوز بدوري أبطال أوروبا سيظل معه "إلى الأبد"	1	7
2	1	إن الفوز بدوري أبطال أوروبا سيظل معه "إلى الأبد"	2	9
3	1	وذلك بعدما ساعده هدفه النادي الملكي على هزيمة أتلتيكو مدريد 1-4 في المباراة النهائية بالبطولة	3	14

Sentences:

id	doc_id	sentence	sentence_order	sentence_length
1	1	قال غاريث بيل، نجم فريق ريال مدريد، إن الفوز بدوري أبطال أوروبا سيظل معه "إلى الأبد"	1	7
2	1	إن الفوز بدوري أبطال أوروبا سيظل معه "إلى الأبد"	2	9
3	1	وذلك بعدما ساعده هدفه النادي الملكي على هزيمة أتلتيكو مدريد 1-4 في المباراة النهائية بالبطولة	3	14

id	doc_id	sentence	sentence_order	sentence_length
1	1	قال غاريث بيل، نجم فريق ريال مدريد، إن الفوز بدوري أبطال أوروبا سيظل معه "إلى الأبد"	1	7
2	1	إن الفوز بدوري أبطال أوروبا سيظل معه "إلى الأبد"	2	9
3	1	وذلك بعدما ساعده هدفه النادي الملكي على هزيمة أتلتيكو مدريد 1-4 في المباراة النهائية بالبطولة	3	14

Words:

Id	Sen_id	Word	Word_order	pos
1	1	قال	1	
2	1	غاريث	2	
3	1	بيل	3	
4	1	نجم	4	
5	1	فريق	5	
6	1	ريال	6	
7	1	مدريد	7	
8	2	إن	1	
9	2	الفوز	2	

Matthew 1

New International Version (NIV)

The Genealogy of Jesus the Messiah

1 This is the genealogy^[a] of Jesus the Messiah^[b] the son of David, the son of Abraham:

2 Abraham was the father of Isaac,
Isaac the father of Jacob,
Jacob the father of Judah and his brothers,
3 Judah the father of Perez and Zerah, whose mother was Tamar,
Perez the father of Hezron,
Hezron the father of Ram,
4 Ram the father of Amminadab,

١كتاب ميلاد يسوع المسيح ابن داود ابن ابراهيم:^٢ ابراهيم ولد إسحاق. وإسحاق ولد يعقوب. ويعقوب ولد يهودا وأخوه.^٣ ويهودا ولد فارص وزارح من ثamar. وفارص ولد حصرؤن. وحضرؤن ولد أرام.^٤ وأرام ولد عمنياداب. وعمنياداب ولد نحشون. ونحشون ولد سلمون.^٥ وسلمون ولد بوعز من راحاب. وبوعز ولد عوبيد من راغوث. وعوبيد ولد يشى.^٦ ويشى ولد داود الملك. وداود الملك ولد سليمان من التي لأوري.^٧ سليمان ولد رحجام. ورحجام ولد أبيا. وأبيا ولد آسا.^٨ وآسا ولد يهوشافاط.

١ كتاب نسب نامه عیسی مسیح بن داود بن ابراهیم: ابراهیم اسحاق را آورد و اسحاق یعقوب را آورد و یعقوب یهودا و برادران او را آورد.^٣ و یهودا، فارص و زارح را از ثamar آورد و فارص، حصرؤن را آورد و حصرؤن، ارام را آورد.^٤ و ارام، عمنياداب را آورد و عمنياداب،



Matthew 1

New International Version (NIV)

The Genealogy of Jesus the Messiah

1 This is the genealogy^[a] of Jesus the Messiah^[b] the son of David, the son of Abraham:

2 Abraham was the father of Isaac,

Isaac the father of Jacob,

Jacob the father of Judah and his brothers,

3 Judah the father of Perez and Zerah, whose mother was Tamar,

Perez the father of Hezron,

Hezron the father of Ram,

4 Ram the father of Amminadab,



^١كتاب ميلاد يسوع المسيح ابن داود ابن إبراهيم: ^٢إِبْرَاهِيمُ وَلَدُ إِسْحَاقَ. وَإِسْحَاقُ وَلَدٌ
يَعْقُوبُ. وَيَعْقُوبُ وَلَدٌ يَهُوذَا وَإِخْوَتُهُ. ^٣وَيَهُوذَا وَلَدٌ فَارِصٌ وَزَارِحٌ مِنْ ثَامَارَ. وَفَارِصٌ
وَلَدٌ حَصْرُونَ. وَحَصْرُونَ وَلَدٌ أَرَامٌ. ^٤وَأَرَامٌ وَلَدٌ عَمِينَادَابٌ. وَعَمِينَادَابٌ وَلَدٌ نَحْشُونَ.
وَنَحْشُونَ وَلَدٌ سَلْمُونٌ. ^٥وَسَلْمُونٌ وَلَدٌ بُوْعَزٌ مِنْ رَاحَابَ. وَبُوْعَزٌ وَلَدٌ عُوبِيدٌ مِنْ رَاعُوتَ.
وَعُوبِيدٌ وَلَدٌ يَسَّى. ^٦وَيَسَّى وَلَدٌ دَاؤُدُ الْمَلَكُ. وَدَاؤُدُ الْمَلَكُ وَلَدٌ سُلَيْمَانٌ مِنْ الَّتِي لَأْوَرِيَا. ^٧وَسُلَيْمَانٌ وَلَدٌ رَحْبَعَامٌ. وَرَحْبَعَامٌ وَلَدٌ أَبِيَا. وَأَبِيَا وَلَدٌ آسَا. ^٨وَآسَا وَلَدٌ يَهُوشَافَاطٌ.

كتاب نسب نامه عیسی مسیح بن داود بن ابراهیم: ^۲ابراهیم اسحاق را آورد و ^۱

اسحاق یعقوب را آورد و یعقوب یهودا و برادران او را آورد. ^۳و یهودا، فارص و زارح را از تamar
آورد و فارص، حصریون را آورد و حصریون، ارام را آورد. ^٤و ارام، عمیناداب را آورد و عمیناداب،

Matthew 1

New International Version (NIV)

The Genealogy of Jesus the Messiah

1 This is the genealogy^[a] of Jesus the Messiah^[b] the son of David, the son of Abraham:

2 Abraham was the father of Isaac,
Isaac the father of Jacob,
Jacob the father of Judah and his brothers,
3 Judah the father of Perez and Zerah, whose mother was Tamar,
Perez the father of Hezron,
Hezron the father of Ram,
4 Ram the father of Amminadab,

¹كتاب ميلاد يسوع المسيح ابن داود ابن إبراهيم: ²إبراهيم ولد إسحاق. وإسحاق ولد يعقوب. ويعقوب ولد يهودا وإخوته. ³ويهودا ولد فارص وزارح من تamar. وفارص ولد حضرن. وحضرن ولد أرام. ⁴وأرام ولد عميناداب. وعميناداب ولد نحشون. ونحشون ولد سلمون. ⁵وسلمون ولد بوعز من راحاب. وبوعز ولد عوبيد من راعوث. وعوبيد ولد يسى. ⁶ويسى ولد داود الملك. وداود الملك ولد سليمان من التي لأوربا. ⁷وسليمان ولد رحهام. ورحهام ولد أبيا. وأبيا ولد آسا. ⁸وآسا ولد يهوشافاط.

¹كتاب نسب نامه عیسی مسیح بن داود بن ابراهیم: ابراهیم اسحاق را آورد و اسحاق یعقوب را آورد و یعقوب یهودا و برادران او را آورد. ³ یهودا، فارص و زارح را از تamar آورد و فارص را آورد و حضرن، ارام را آورد. ⁴ ارام، عمیناداب را آورد و عمیناداب،



English	Arabic	Persian
genealogy	كتاب	كتاب
Jesus	المسيح	مسيح
David	داود	داود
Abraham	إبراهيم	ابراهيم
Isaac	إسحاق	اسحاق
Jacob	يعقوب	يعقوب
Judah	يهودا	يهودا
Hezron	حصرون	حصرون
Perez	فارص	فارص



How

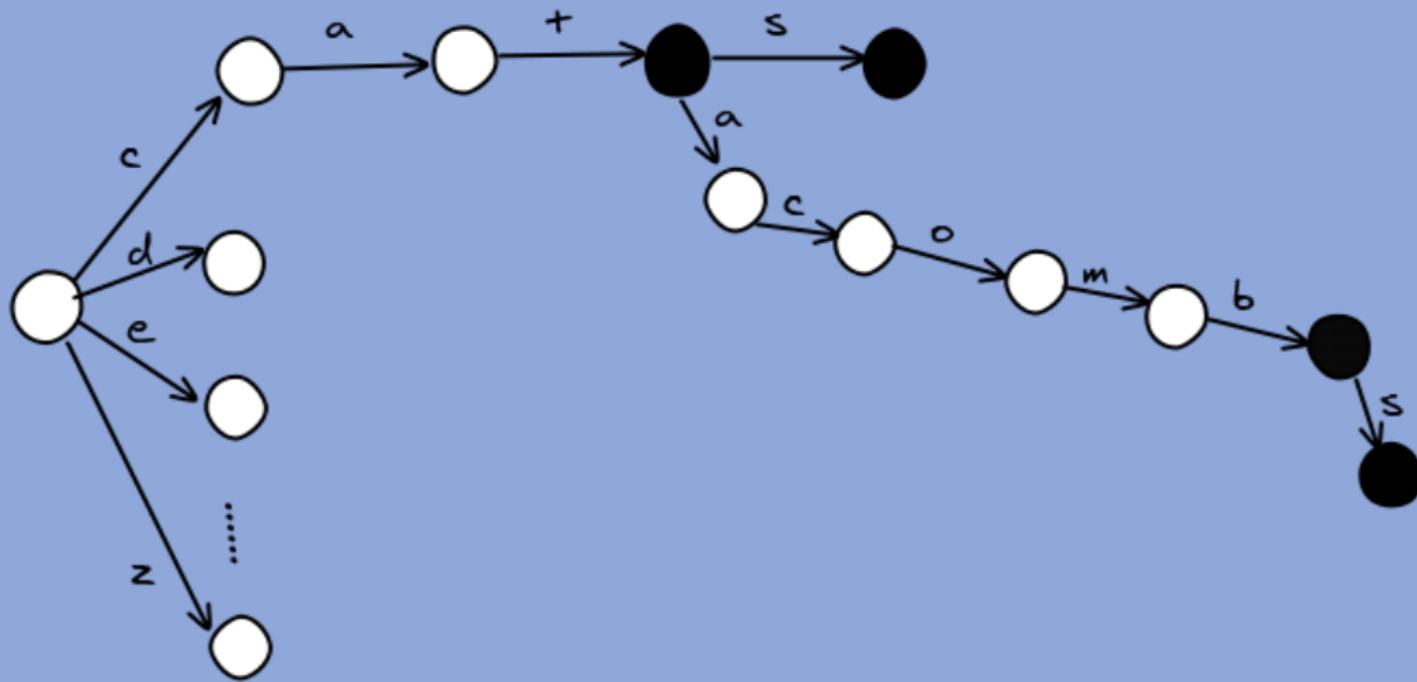
- I co-occurrence: 4L-4R?
- I part of speech tags: wiktionary, University of Tehran, Uni Leipzig
- I Distance metrics: Levenshtein using a trie
- I Machine learning algorithm

و چون عیسی در ایام هیرودیس پادشاه در بیتلحم بهودیه تولد یافت، ناگاه مجوسي چند از مشرق به اورشلیم آمد، گفتند: «جاست آن مولود که پادشاه بهود است زیرا که ستاره او را در مشرق دیده ایم و برای پرستش او آمده ایم»

وَ لَمَّاْ وُلِدَ يَسُوعُ فِي بَيْتِ لَحْمِ الْيَهُودِيَّةِ، فِي أَيَّامِ هِيرُودِيسِ الْمَلِكِ، إِذَا مَحْوَسٌ مِّنَ الْمَشْرُقِ قَدْ جَاءُوا إِلَى أُورْشَلِيمَ 2فَائِلِينَ: «أَيْنَ هُوَ الْمَوْلُودُ مَلِكُ الْيَهُودِ؟ فَإِنَّا رَأَيْنَا نَجْمَةً فِي الْمَشْرُقِ وَأَتَيْنَا لِتَسْجُدَ لَهُ».«

- I Trie-Join: Efficient Trie-based String Similarity Joins with Edit-Distance Constraints, Jiannan Wang et al. <http://dl.acm.org/citation.cfm?id=1920992>
- I 4L-4R: Distributional Semantics and Automatic Semantic Information Extraction: An Introduction

Trie



 <http://stevehanov.ca/blog/index.php?id=114>

results

و جون عیسی در آیام هیرودیس پادشاه در بیتلحم یهودیه تولد یافت ، ناگاه محسوسی چند از مشرق به اورشلیم آمده ، گفتند: « جاست آن مولود که پادشاه یهود است زیرا که ستاره او را در مشرق دیده ایم و برای پرستش او آمده ایم »

وَ لَمَّا وُلِدَ يَسُوعُ فِي بَيْتِ لَحْمِ الْيَهُودِيَّةِ، فِي أَيَّامِ هِيرُودِسَ الْمَلِكِ، إِذَا مَحْوَسٌ مِنَ الْمَشْرُقِ قَدْ جَاءَهُوا إِلَى أُورْشَلِيمَ 2 قَائِلِينَ: «أَيْنَ هُوَ الْمَوْلُودُ مَلِكُ الْيَهُودِ؟ فَإِنَّا رَأَيْنَا نَجْمَةً فِي الْمَشْرُقِ وَأَتَيْنَا لِنَسْجُدُ لَهُ».»

	Bible(Chapter1) (507)	Bible (2+3+4) (1619)	UN charter (11753)
euclidean distance	96.078093	180.174915	1154.207954
manhattan distance	763	2449	19698
hamming distance	0.982968	0.990393	0.995295
sokalmichener distance	0.985185	0.9903	0.986563

و چون عیسی در **ایام هیرودیس** پادشاه در بیتلحم یهودیه **تولد** یافت ، ناگاه **محوسی** چند از **شرق** به **اورشلیم** آمده ، گفتند: « جاست آن **مولود** که پادشاه **یهود** است زیرا که ستاره او را در **شرق** دیده ایم و برای پرستش او آمده ایم »

وَ لَمَّا وُلِدَ يَسُوعُ فِي بَيْت لَحْم الْيَهُودِيَّةِ، فِي أَيَّام هِيرُودِيس الْمَلِكِ، إِذَا مَحْوَسٌ مِنَ الْمَشْرُقِ قَدْ جَاءُوا إِلَى أُورْشَلِيمَ 2 قَائِلِينَ: «أَيْنَ هُوَ الْمَوْلُودُ مَلِكُ الْيَهُودِ؟ فَإِنَّا رَأَيْنَا نَجْمَةً فِي الْمَشْرُقِ وَأَتَيْنَا لِنَسْجُدَ لَهُ».

	Bible(Chapter1) (507)	Bible (2+3+4) (1619)	UN charter (11753)
euclidean distance	96.078093	180.174915	1154.207954
manhattan distance	763	2449	19698
hamming distance	0.982968	0.990393	0.995295
sokalmichener distance	0.985185	0.9903	0.986563



و جون عیسی در **ایام هیرودیس** پادشاه در بیتلحم یهودیه **تولد** یافت ، ناگاه **محوسی** چند از **شرق** به **اورشلیم** آمده ، گفتند: « جاست آن **مولود** که پادشاه **یهود** است زیرا که ستاره او را در **شرق** دیده ایم و برای پرستش او آمده ایم »

وَ لَمَّا وُلِدَ يَسُوعُ فِي بَيْتِ لَحْمِ الْيَهُودِيَّةِ، فِي أَيَّامِ هِيرُودِيسَ الْمَلِكِ، إِذَا مَحْوَسٌ مِنَ الْمَشْرُقِ قَدْ جَاءُوا إِلَى أُورْشَلِيمَ 2 قائلین: «أَيْنَ هُوَ الْمَوْلُودُ مَلِكُ الْيَهُودِ؟ فَإِنَّا رَأَيْنَا نَجْمَةً فِي الْمَشْرُقِ وَأَتَيْنَا لِنَسْجُدَ لَهُ».

	Bible(Chapter1) (507)	Bible (2+3+4) (1619)	UN charter (11753)
euclidean distance	96.078093	180.174915	1154.207954
manhattan distance	763	2449	19698
hamming distance	0.982968	0.990393	0.995295
sokalmichener distance	0.985185	0.9903	0.986563

Perspective

IF

I finish it up to the end of Vorlesungszeit

I get more resources, corpora

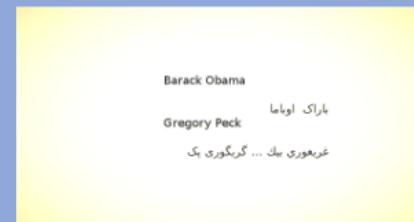
THEN

I (making translator)

I Latinyser

ELSE

I keep it small



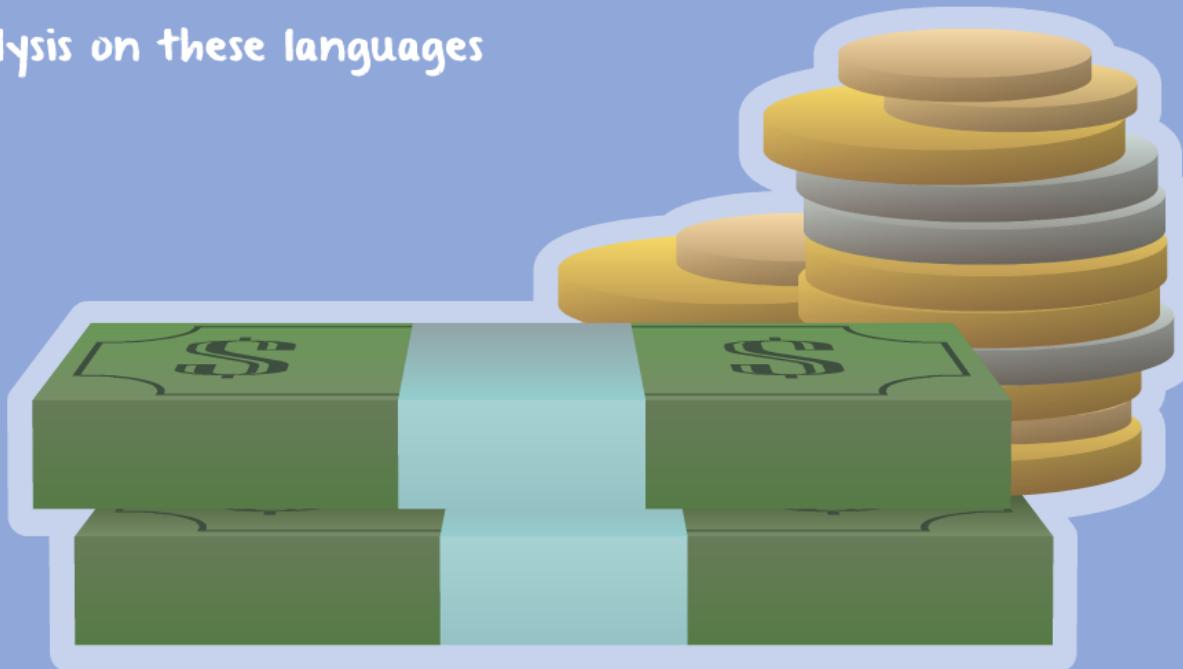
Barack Obama

باراک اوباما

Gregory Peck

غريغوري بيك ... گريگوري پك

- New and trustable analysis on these languages
- Latinyser





Jiannan Wang Jianhua Feng Guoliang Li,

Trie-Join: Efficient Trie-based String Similarity Joins with Edit-Distance Constraints.

Proceedings of the VLDB Endowment VLDB Endowment Homepage archive Volume 3 Issue 1-2,
September 2010, Pages 1219-1230



Steve Hanov

Fast and Easy Levenshtein distance using a Trie

<http://stevehanov.ca/blog/index.php?id=114>



SQLite: The Definitive Guide to SQLite (Expert's Voice in Open Source)

<http://www.amazon.com/gp/product/1430232250>

<http://www.tutorialspoint.com/sqlite/>



POS: http://en.wiktionary.org/wiki/Category:Persian_parts_of_speech

http://en.wiktionary.org/wiki/Category:Arabic_parts_of_speech



Database Research Group

ECE Department, University of Tehran

<http://ece.ut.ac.ir/DBRG/>





keep
going

Analyzing and comparing
Arabic and Persian

Tariq and Iman

