

Workshop on Arabic Corpus Linguistics

11-12 April 2011, Lancaster University

Corpus linguistics resources and tools for Arabic lexicography

Majdi Sawalha and Eric Atwell

School of Computing,

University of Leeds, Leeds, LS2 9JT, UK

<http://www.comp.leeds.ac.uk/sawalha>

<http://www.comp.leeds.ac.uk/eric>

Outline

- Introduction
 - Oxford University Dictionaries
 - Monolingual Dictionaries
 - Bilingual Dictionaries
 - Arabic & Arabic NLP
 - Difficulties
 - Morphological analysis
 - Traditional Arabic lexicography
 - Constructing the Arabic broad-coverage lexical resource
 - Key notes
 - Conclusions
 - References

Oxford Dictionaries

- Searching for a word



- Dictionary entries are the lemmas of the word.
- Lemmatising of input words is needed to direct the search for the input word to the correct dictionary entry.
- What happens if the user entered a miss-spelled (unknown) word?



- Gives suggestions of similar words

No exact results found for "university" in the dictionary.

Did you mean *university*?

Did you mean *inveracity*?

Did you mean *intercity*?

university

inveracity

intercity

Dictionary Entry


- Dictionary entry: Information provided

[Dictionary/Thesaurus](#) | [Example sentences](#) | [For writers and editors](#) | [Writing skills](#) | [Puzzles](#)

[Print](#) | [Save](#) | [Email](#) | [Cite](#)

Dictionary Entry

university

Pronunciation: /juːnɪˈvɜːsɪti/ 

noun (plural universities)

a high-level educational institution in which students study for degrees and academic research is done:

I went to university at the Sorbonne
his daughter is at university



Phrases

the university of life
the experience of life regarded as a means of instruction

– *That is acquiring the knowledge which you will do well to put into practice as you travel to the next stage of your journey through the university of life.*

Origin:
Middle English: from Old French *universite*, from Latin *universitas* 'the whole', in late Latin 'society, guild', from *universus* (see UNIVERSE)

12/4/2011

Text size:  

Line-break: [On](#) | [Off](#)

Pronunciation

POS + Plural form

Meaning

Examples

Phrases

Meaning

Examples

Origin

'university' in Oxford Dictionaries Online
[View thesaurus entry](#)
[Full-text search](#)
[See all example sentences](#)

'university' in the OED Online
[See the full history](#)

'university' in ORO Online
[Search references](#)


Look for translations of 'university' in
[Chinese](#) | [French](#) | [German](#) | [Italian](#) | [Russian](#) | [Spanish](#)

Other Dictionaries

My entries(1)

Dictionary entries

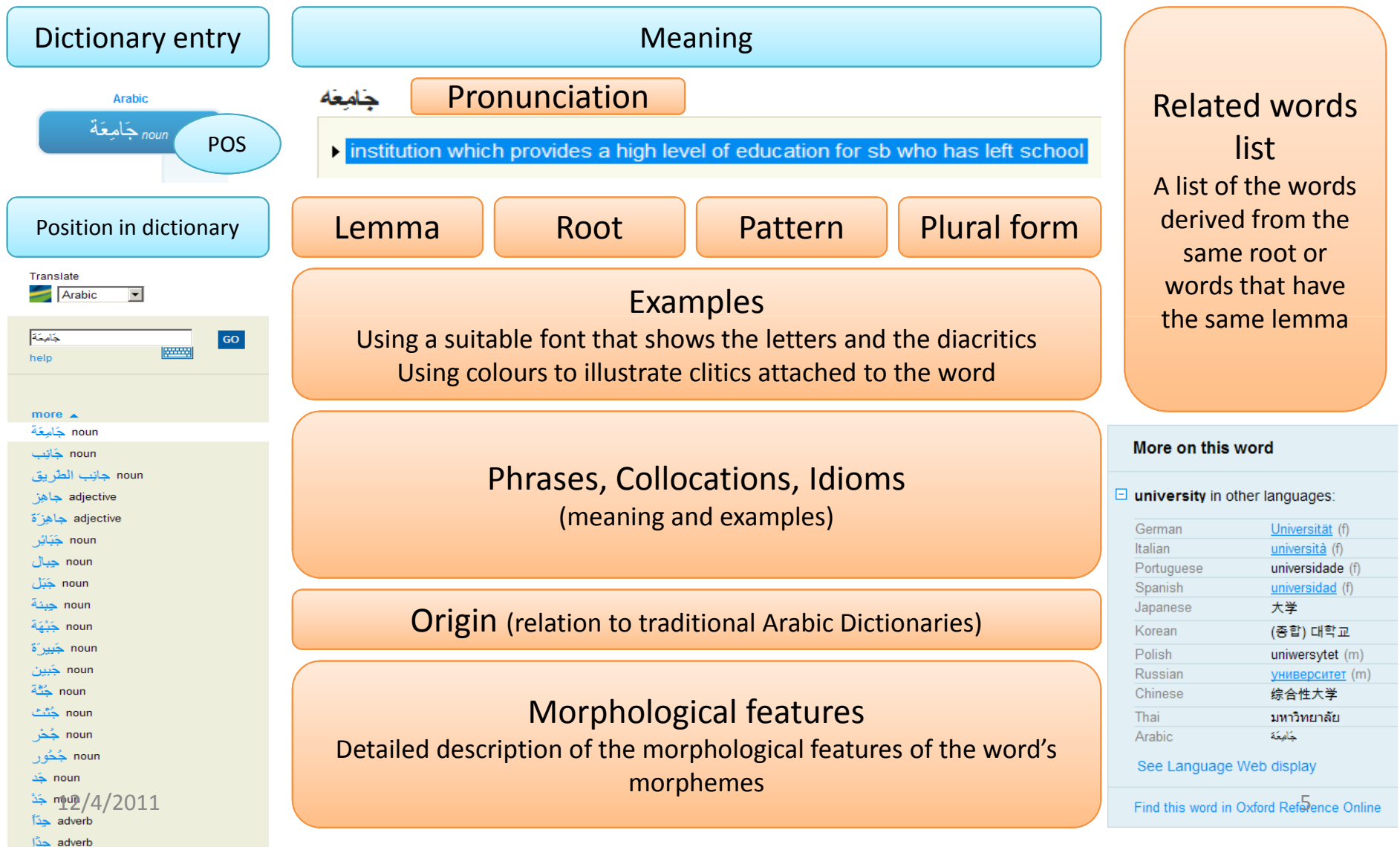
Jump to:



Entry
universally
Universal Postal Union
Universal Product Code
universal quantifier
universal recipient
universal set
universal suffrage
Universal Time
universe
university
University City
University of the Th...
University Park
univocal
univocality
univocally
Unix
unjoin
unjoined
unjoint

4

What information can be provided by the Arabic dictionaries?



'university' in OED online

Oxford English Dictionary

Lost for Words? Find Word

SECOND EDITION 1989

university, *n.*

Pronunciation Spellings Etymology Quotations Date chart Additions

(juːnɪˈvɜːsɪti) Forms: 6 **vniuersite**, 5 **-versite**, 5-6 **-uersitee**, 4-5 **vnyuersite(e)**, 5-6 **-uersyte(e)**; 5-7 **vniuersitie** (6 *Sc. wni-*), 6 **-tye**, **vnyuersytye**, **-tie**, 7 **vniuersity**, **-versity**, 6-7 **universitie**, 7- **university**. See also [varsity](#), [versity](#). [a. AF. *université*, *universeté*, *univercyté*, OF. *universitei*, *universiteit*, *université* (13th c.; mod.F. *université*, = Pr. *universitat*, It. *università*, Sp. *universidad*, Pg. *-idade*; also in sense 1 MDu. *universitet*, MDu. and Du. *universiteit*, MG., MLG. *universitete*, MHG. *universitêt*, G. *universität*, Dan., Sw. *universitet*):—L. *ūniversitat-*, *ūniversitas*, (1) the whole, entire number, universe, (2) in later and mediæval Latin (chiefly in legal use), a society, company, corporation, or community regarded collectively; f. L. *ūniversus* (see [UNIVERSE](#)).]

I. 1. a. The whole body of teachers and scholars engaged, at a particular place, in giving and receiving instruction in the higher branches of learning; such persons associated together as a society or corporate body, with definite organization and acknowledged powers and privileges (esp. that of conferring degrees), and forming an institution for the promotion of education in the higher or more important branches of learning; also, the colleges, buildings, etc., belonging to such a body. In recent use, const. without article: **at** (or **to**) **university**, etc.
Sometimes, especially in former use, synonymous with *college*: see [COLLEGE](#) *n.* 4c.

c1384 c1300 c1400 1579 1509 1661 1644 1785 1725 1702 1868 1856 1840 1968 1969

c1300 *St. Edmund in S. Eng. Leg.* I. 438/256 So þat he bigan at Oxenford of diuinite, So noble alosed þer nas non in al þe vniuersite. *Ibid.* 439/278 He bigan so deope desputi of þe trinite, þat gret wonder me hadde þurf al þe vniuersite. **c1384** ? *WYCLIF Wks.* (1880) 157 Heþene mennus lawis and worldly clerks statutis ben red in vnyuersitees. **c1400** *Rom. Rose* 6769 At Parys..he had..The accorde of the vniuersite, And of the puple as semeth me. **c1425-** [see [COLLEGE](#) *n.* 4]. **c1450** *Godstow Reg.* 438 The house..that Robert of Staunton held of the vnyuersite of Oxenford. **1509** *FISHER Funeral Sermon. Ctess Richmond Wks.* (1876) 301 The studentes of bothe the vnyuersytees. **1579** *W. WILKINSON Confut. Fam. Love* 40 They labour to put out the eyes of this land (the Vniuersities I meane). **1644** *MILTON Educ.* 3 This place should be at once both School and University. **1661** *LAMPLUGH in Extr. St. Papers Friends Ser.* II. (1911) 126 University, Town and Country are far more active and vigilant then before. **1702** *LUTTRELL Brief Rel.* (1857) V. 145 A patent..for founding an university.., to be called king Williams university. **1725** *BAILEY Erasm. Colloq.* (1733) 259 Are you going to Louvain to see the University? **1785** *J. ADAMS Wks.* (1854) IX. 530 He is anxious to study some time at your university [= Harvard College] before he begins the study of law. **1840** *CARLYLE Heroes v.* (1858) 305 Universities are a notable, respectable product of the modern ages. **1856** *STANLEY Sinai & Pal.* x. 364 The great Jewish university

12/11/2011 List by entry List by date Entry map Simple search Advanced search

Print Mail Help Sign out

Subscriber: University of Leeds

OXFORD UNIVERSITY PRESS

Traditional Arabic Lexicography

- Arabic lexicography is one of the original and deep-rooted arts of Arabic literature.
- The first lexicon constructed was *kitāb al-‘ayn* كتاب العين *al-‘ayn lexicon* by *al-farāhīdī* (died in 791).
- Over the past 1200 years, many different kinds of Arabic language lexicons were constructed; these lexicons are different in ordering, size and aim or goal of construction.
- Many Arabic language linguists and lexicographers studied the construction, development and the different methodologies used to construct these lexicons.

Ordering lexical entries in the Arabic lexicons

- *al-ḥalīl* **methodology** [5 lexicons]
 - Listed the lexical entries based on the pronunciation of the letters starting from the farthest in the mouth to the nearest
- *abī‘ubayd* **methodology** [3 lexicons]
 - Listed the lexical entries based on similarity in meaning.
- *al-ğawharī* **methodology** [4 lexicons]
 - Listed the lexical entries based on last letter of the word.
- *al-barmakī* **methodology** [11 lexicons]
 - Listed the lexical entries alphabetically.
- **Modern dictionaries** uses a combination of root/word as lexical entries arranged alphabetically.

Search for the meaning of any Arabic root

Enter Arabic Root (أدخل الجذر العربي): جمع search

الجذر The Root

جمع

تاج العروس من جواهر القاموس *tag al-'arūs min ġawāhir al-qāmūs*

جمع : الجمعُ ، كالمَنعِ : تَأْلِيفُ الْمُتَفَرِّقِ . وفي الْمُرَدَّاتِ لِلرَّاغِبِ وَبَعَهُ الْمُصَنَّفُ فِي الْبَصَائِرِ : الْجَمْعُ : ضَمُّ الشَّيْءِ بِتَقْرِيبِ بَعْضِهِ مِنْ بَعْضٍ . يُقَالُ : جَمَعْتُهُ فَاجْتَمَعَ . وَالْجَمْعُ أَيْضاً : الدُّقْلُ . يُقَالُ : مَا أَكْثَرَ الْجَمْعَ فِي أَرْضِ بَنِي فُلَانٍ ، أَوْ هُوَ صِنْفٌ مِنَ الثَّمَرِ مُخْتَلِطٌ مِنْ أَنْوَاعٍ مُتَفَرِّقَةٍ ، وَلَيْسَ مَرْغُوباً فِيهِ ، وَمَا يُخْلَطُ إِلَّا لِرَدَائِهِ . وَمِنْهُ الْحَدِيثُ : بَعِ الْجَمْعَ بِالْدَّرَاهِمِ ، وَابْتِغِ بِالْدَّرَاهِمِ جَنِيحاً . أَوْ هُوَ النَّخْلُ خَرَجَ مِنَ النَّوَى لَا يُعْرَفُ اسْمُهُ . وَقَالَ الْأَصْمَعِيُّ : كُلُّ لَوْنٍ مِنَ النَّخْلِ لَا يُعْرَفُ اسْمُهُ فَهُوَ جَمْعٌ . وَقَالَ ابْنُ دُرَيْدٍ : يَوْمُ الْجَمْعِ : يَوْمُ الْقِيَامَةِ . وَقَالَ ابْنُ عَبَّادٍ : الْجَمْعُ : الصَّمْعُ الْأَحْمَرُ . وَالْجَمْعُ : جَمَاعَةُ النَّاسِ ، ج : جُمُوعٌ ، كَبْرُوقٌ وَبُرُوقٌ ، كَالْجَمِيعِ ، كَمَا فِي الْعُبَابِ . وفي الصَّحاحِ : الْجَمْعُ قَدْ يَكُونُ مَصْدَرًا ، وَقَدْ يَكُونُ اسْمًا لَجَمَاعَةِ النَّاسِ ، وَيُجْمَعُ عَلَى جُمُوعٍ ، زَادَ فِي اللِّسَانِ : وَالْجَمَاعَةُ ، وَالْجَمِيعُ ، وَالْمَجْمَعُ ، كَالْجَمْعِ ، وَقَدْ اسْتَعْمَلُوا ذَلِكَ فِي غَيْرِ النَّاسِ حَتَّى قَالُوا : جَمَاعَةُ الشَّجَرِ ، وَجَمَاعَةُ النَّبَاتِ . وَالْجَمْعُ : لَبَنٌ كُلُّ مَضْرُورَةٍ ، وَالْفَوَاقِ : لَبَنٌ كُلُّ بَاهِلَةٍ ، وَسَيَّاتِي فِي مَوْضِعِهِ ، وَإِنَّمَا ذُكِرَ هُنَا اسْتِطْرَادًا ، كَالْجَمِيعِ . وَجَمْعٌ بِلَا لَامٍ : الْمَزْدَلْفَةُ ، مَعْرِفَةٌ ، كَعَرَفَاتٍ ، لَااجْتِمَاعِ النَّاسِ بِهَا ، وفي الصَّحاحِ : فِيهَا . وَقَالَ غَيْرُهُ : لِأَنَّ آدَمَ وَحَوَّاءَ لَمَّا أَهْبِطَا اجْتَمَعَا بِهَا . قَالَ أَبُو ذُوئَيْبٍ : % (قَبَاتٍ بِجَمْعٍ ، ثُمَّ تَمَّ إِلَى مِثْنٍ % فَأَصْبَحَ رَادًّا يَتَّبِعِي الْمَرْجَ بِالسَّحْلِ) % وَقَالَ ابْنُ دُرَيْدٍ : يَوْمُ جَمْعٍ : يَوْمُ عَرَفَةَ ، وَأَيَّامُ جَمْعٍ : أَيَّامُ مِثْنٍ . وَالْمَجْمُوعُ : مَا جُمِعَ مِنْ هَا هُنَا وَهَا هُنَا ، وَإِنْ لَمْ يُجْعَلْ كَالشَّيْءِ الْوَاحِدِ نَقَلَهُ الْجَوْهَرِيُّ وَالصَّاغَانِيُّ وَصَاحِبُ اللِّسَانِ . وَالْجَمِيعُ : ضِدُّ الْمُتَفَرِّقِ ، قَالَ قَيْسُ بْنُ ذَرِيحٍ : % (فَقَدْتُكَ مِنْ نَفْسٍ شَعَاعٍ ، فَإِنِّي % تُهَيْثُكَ عَنْ هَذَا ، وَأَنْتَ جَمِيعٌ) % وَالْجَمِيعُ : الْجَيْشُ . قَالَ لَبِيدٌ رَضِيَ اللَّهُ عَنْهُ : (فِي جَمِيعِ حَافِظِي عَوْرَاتِهِمْ % لَا يَهْمُونَ بِإِدْعَاكِ الشَّلَلِ) %) وَالْجَمِيعُ الْحَيُّ الْمُجْتَمِعُ . قَالَ لَبِيدٌ رَضِيَ اللَّهُ عَنْهُ يَصِفُ الدِّيَارَ : % (عَرِيتُ ، وَكَانَ بِهَا الْجَمِيعُ فَأَبْكُرُوا مِنْهَا ، فَعُودِرَ نُؤْيُهَا وَتُمَامُهَا) % وَجَمِيعٌ : عِلْمٌ ، كَجَمَاعٍ ، وَهَمَّا كَثِيرَانِ فِي الْأَعْلَامِ . وفي الصَّحاحِ وَالْعُبَابِ : أَتَانُ جَمَاعٌ : إِذَا حَمَلَتْ أَوَّلَ مَا تَحْمِلُ . وَقَالَ ابْنُ شُمَيْلٍ : جَمَلُ % جَمَاعٌ ، وَنَاقَةٌ ، جَمَاعَةٌ ، إِذَا أَخْلَفَا بُزُولًا ، قَالَ : وَلَا يُقَالُ هَذَا إِلَّا بَعْدَ أَرْبَعِ سِنِينَ . هَكَذَا فِي النَّسَخِ ، وَصَوَّاهُ عَلَى مَا فِي الْعُبَابِ وَالتَّكْمِلَةِ : وَلَا يُقَالُ هَذَا بَعْدَ أَرْبَعِ سِنِينَ ، مِنْ غَيْرِ حَرْفٍ إِلَّا اسْتِثْنَاءً . إِذَا كَانَتْ تَصْلُحُ لِلْإِكَافِ وَالسَّرْجِ ، نَقَلَهُ الصَّاغَانِيُّ . وَقَدَّرُ جَمَاعٌ ، وَجَمَاعَةٌ ، وَجَمَاعٌ ، كَكِتَابٍ ، أَيْ عَظِيمَةٍ ، ذَكَرَ الصَّاغَانِيُّ الْأَوَّلَ 4/2011 وَالثَّانِيَةَ . وَاقْتَصَرَ الْجَوْهَرِيُّ عَلَى الثَّانِيَةِ . وَنَسَبَ صَاحِبُ اللِّسَانِ الْأَخِيرَةَ إِلَى الْكَسَائِيِّ . قَالَ الْكَسَائِيُّ : أَكْبَرُ الْبِرَامِ الْجَمَاعُ ، ثُمَّ التَّتِي تَلِيهَا الْمُتَكَلَّةُ . وَقِيلَ : قَدَّرُ جَمَاعٌ وَجَمَاعَةٌ : هِيَ الَّتِي

The use of Corpora in building dictionaries – Example 1

Enter a word or phrase to see examples

Showing 1-20 of 4513 example sentences for “university”

- 1 Two days later I'm told, I get a call from the state university, and the state university says, we have a problem.
- 2 He began his university studies in 1951 when he entered Tokyo University.
- 3 By education he means at the university of life as well as Le Havre University, where he got his degree in mathematics, giving himself a safety net in case he does tumble from the current pedestal.
- 4 This 70-something volunteer at the University of the Third Age in Clifford Street, York, which organises courses and activities for the over-50s, thought her life was heading for the slow lane when she turned 50.
- 5 He was born the son of a poor mill owner, and was the first in his family to go to university.
- 6 Get out and enjoy the artistic side of our often misrepresented university.
- 7 If you join a course run by the University of the Third Age do not ask your fellow student agriculture
- 8 That is acquiring the knowledge which you will do well to put into practice as you travel agriculture
- 9 The contention that the First Amendment forbids a state university to provide a venue agriculture
Christian beliefs is absurd. agriculture
- 10 His holidays from university have found him up on the release site, near Pewsey, working agriculture
- 11 The undergraduate library at my alma mater - a state university with about 15,000 users agriculture
- 12 The next meeting of the Evesham branch of the University of the Third Age will be held agriculture
- 13 I didn't have the benefit of a great education as a young man, but the university of life. agriculture
- 14 The clearest examples of its proponents were on the religious studies faculty at the state agriculture
- 15 This is alleged to be the first time in the 800 year history of that university. agriculture
- 16 The undergraduate American students were majoring in education at a large state university agriculture
- 17 She didn't go into higher education, leaving school at 16, but says, with a wink that acknowledges agriculture
university of life. agriculture

Selecting examples
from concordance lines

Lexicographers

crops specialists at their state land-grant university to obtain the best information for their Department of Horticulture, Michigan State University, East Lansing, MI </p><p> Jim Motes, Department of Horticulture, Oklahoma State University, Stillwater, OK </p><p> Sam Walker, Jersey Extension Specialist in Vegetables at Rutgers University in New Brunswick, New Jersey, have devoted , and Jersey Gem . </p><p> Michigan State University Asparagus Cultivar Yield Trials </p><p> Locations First harvest in 1994 . Oklahoma State University Bixby Vegetable Research Station </p><p> Pathologist located at your state land-grant university to provide seed treatment information . developed by Dr. Stephen Garrison of Rutgers University . Grooves on each side of the plateau of diseases . </p><p> Studies at Michigan State University have shown over a nine-year harvest period benefit . </p><p> Results of a Michigan State University study (20) indicate that pre-plant applications yield . </p><p> Studies done at Michigan State University (22) showed that after six years of trickle-irrigating with data available from USDA agencies and university and industry drainage specialists . Drainage multi-institutional project lead by the University of Florida and funded by the United Soybean conducted cooperatively with The Ohio State University at the Wooster Branch Station of OARDC Research and Development Center and Ohio State University Extension, The Ohio State University ; State University Extension, The Ohio State University ; USDA-ARS Soil Drainage Research Unit Erie Protection Fund, OARDC, Ohio State University Extension, Ohio Sea Grant Program, North and analyzed in the laboratory at Purdue University . </p><p> Small amounts of pesticide move watershed to determine the extent to which University of Illinois fertilizer recommendations impacts of drainage has been developed by the University of Minnesota and scientists within federal

12/4/2011

Lemmatized Arabic Internet Corpus

<http://corpus.leeds.ac.uk/query-ar.html>

Querying Arabic Corpora

☐ Arabic Internet
 ☐ Al Hayat News
 ☐ Arabic Wikipedia
 ☐ Corpus of Contemporary Arabic
 ☐ Computer Science corpus
 ☐ Arabic legal texts, v2

☐ CQP syntax only ([Examples](#))
 [Click here for getting help on the query interface](#)

Set parameters of your query

Concordance

Context: (c for characters, w for words)

Sort by: ☐ Document ☐ Frequency ☒ lemma ☐ word

Then by: ☐ left ☒ right

Output: lines

Collocations

Collocation scores: ☐ Mutual Information ☐ T-score ☒ Loglikelihood

Context: words on the left words on the right

POS tag of the collocate: POS tags

I also created the lists of the most frequent word forms in [Internet](#), [LDC](#), [Wiki](#). After lemmatisation done by [Majdi Sawalha](#) there is also the frequency list of

The corpora are:

1. The Internet corpus was compiled using the procedure described in r
2. The Al Hayat corpus — from Al Hayat data (1999-2001) compiled t
3. The Wikipedia corpus — from the public Wiki data retrieved on July
4. CCA corpus — from [Latifa Al-Sulaiti](#).
5. The Arabic Legal Corpus — from keywords collected by Hanem El-
6. Computer Science corpus of Arabic — from keywords collected by:

The interface was developed by Serge Sharoff; contact me at s.sharoff@leeds.ac.uk

[Back to the query window](#)

See 100 examples of '[word="جامعة"] cut 100' in I-AR-LEMMA

- الفصوص ص 16 & quot; والقلم في نظر الفارابي ، هو ملك روحاني ! جامعة لها علاقة بينها وبين نفسها ، ولا تكون وحدانيته وحدانية ، أو بتعبير آخر & quot; كما ذكرنا في الباب الثاني & quot; جامعة فمن المؤكد أن تكون ذاته عينها مع وحدانية جوهرها ، هي بنفسها الفصل الثاني توافق الوحدانية الجامعة المانعة مع وحدانية . جامعة يمكن أن تكون وحدانية مجردة أو مطلقة ، بل وحدانية شاملة أو وقال أيضاً: إن كان عقله ذاته ، فيرجع الكل إلى ذاته ، فلا . جامعة على تميزه بكثرة ، أو بتعبير آخر على أن وحدانيته هي وحدانية ويتفق معنا على ذلك كثير من علماء الدين ، فمثلاً قال صاحب . جامعة إذا ليس <UNDEF/تعبير> في القول إن وحدانيته هي وحدانية ، أو جامعة مانعة. إنما كانوا يدعونها الوحدانية المطلقة كما ولا يسمح للمدرسات الشيعيات بالاعتماد على <STOP/كما> ، جامعة مواد الدين ولا أن يحجب مديرات مدارس ولا وجهات ولا مدرسات يرجع إلى أن الإسلام يسند إلى الله صفات وأسماء كثيرة ، هي ، جامعة بأن وحدانية الله هي <مدرسة/مدرسة> بتعبير آخر وحدانية أبو الوفا الأهلية ماجستير في الأدب وتحضر الدكتوراة تقول لي جامعة أعرفك إنها الأستاذة خولة علم الدين. أستاذة النحو في - أم القرى - فرع الطائف قبل أن تستقل الطائف بجامعة خاصة بها جامعة شخصية في المكتب الثقافي السعودي بالقاهرة بخصوص العمل في أو شاملة ، أن هذه الجامعية أو الشمول هي ذاته وصفاته. لكن جامعة يعتقد بعض الفلاسفة الذين أدركوا أن وحدانية الله هي وحدانية افتراضية سورية ، أي تعليمًا عن بعد. وتتوحد دلالة هذه جامعة يملكون مؤهلاته (التعليم الموازي) ، ومكتب الإنترنت من إنشاء الأزهر. وعلاء ديب والمرشح عبد الهندي ، ومجموعة من عائلة جلس جامعة عضو مجلس طلبة & apos; والمرشح محمد الغول & apos; فاروق جلس النيابة العامة فيما قرر المتهم عبد الكريم <UNDEF> الأزهر جامعة الدعوى لتحصل فيما أبلغت به كلية الشريعة والقانون بدمهور ، الأزهر ، والأخ طارق الهندي ، والأخ الدكتور إيهاب الأغا جامعة بجامعة الأقصى ، والمهندس سعد الدين خرما أمين سر مجلس أمناء الأزهر ، وعضو عدد من الجمعيات العلمية والأدبية ، وصاحب جامعة - حول الرواية ، وهو أستاذ للأدب والنقد كلية اللغة العربية الأزهر على النحو التالي: من له سلطة إحالة الطلاب للتأديب جامعة بهذه اللاتحة. وجاء فيها التنظيم القانوني لتأديب طلاب 1975 / الأزهر وسياسية الفصل العنصري جاءت في العدد 1014 الصادر في 11 جامعة المتمدن ويتضح ذلك من صدر كل مقالة منسوبة للمتهم فمقالة الإمارات العربية بالعين في العام الدراسي 1977 - 1978 تخرج جامعة زاد عدد التلاميذ زيادة كبيرة بالمدارس الابتدائية. وأفتحت

12/4/2011

The use of Corpora in building dictionaries – Example 2

university

noun

1

college, academy, educational establishment/institution, institute, varsity

- *North American school*
- *historical polytechnic*

[View dictionary entry »](#)

Headword Match (1) Synonym Match (5) Antonym Match (0)

Showing 1-5 of 5 results for "university"

Sort by: [A to Z](#) | [Relevance](#)

academy

...educational institution, training establishment, centre of learning, school, college, **university**, institute, seminary, conservatory, conservatoire, *historical polytechnic*...

[Go to Thesaurus entry »](#) [Go to Dictionary entry »](#)

college

...educational institution, training establishment, centre of learning, seat of learning, school, academy, **university**, institute, seminary, conservatory, conservatoire, *historical polytechnic*...

[Go to Thesaurus entry »](#) [Go to Dictionary entry »](#)

academic

...educational, scholastic, instructional, pedagogical, school, college, collegiate, **university**...

[Go to Thesaurus entry »](#) [Go to Dictionary entry »](#)

institute

...organization, establishment, institution, foundation, centre, academy, school, college, **university**, conservatory, seminary, centre of learning, seat of learning, society, association...

[Go to Thesaurus entry »](#) [Go to Dictionary entry »](#)

institution

...organization, establishment, institute, foundation, centre, academy, school, college, **university**, conservatory, seminary, centre of learning, seat of learning, society, association...

[Go to Thesaurus entry »](#) [Go to Dictionary entry »](#)

12/4/2011

university new_oec_plus_biweb_alt freq = 198767

Lemma	Score	Freq
college	0.504	207476
institution	0.416	190572
school	0.41	981325
University	0.371	461863
organization	0.34	416383
agency	0.339	215584
hospital	0.326	276588
department	0.324	203022
education	0.32	357278
community	0.319	672366
club	0.313	360545
council	0.312	263624
church	0.31	279459
firm	0.309	212234
industry	0.302	434801
business	0.298	845818
union	0.295	223911
society	0.293	346017
government	0.293	1199449
company	0.293	1185039
bank	0.291	278006
student	0.287	775851
authority	0.287	347240
sector	0.284	197731

The use of Corpora in building dictionaries – Example 2, cont.

university/college new_oec_plus_biweec_alt freq = 198767/207476

Common patterns

- Frequency lists /measurements can compare between two words, and find how these words are related to each other.

university	6.0	4.0	2.0	0	-2.0	-4.0	-6.0	college						
X at-i N*	16968	10006	11.7	6.4	N* 's N	7439	3212	8.5	3.4	X out_of-i N*	421	1325	2.8	8.3
study	<u>1055</u>	<u>339</u>	7.0	5.3	campus	<u>176</u>	<u>62</u>	5.7	4.2	drop	<u>99</u>	<u>345</u>	3.7	5.4
tuition	<u>59</u>	<u>109</u>	5.3	6.4	endowment	<u>24</u>	<u>17</u>	5.3	5.2	fresh	<u>24</u>	<u>139</u>	2.4	4.9
lecture	<u>80</u>	<u>47</u>	6.4	6.0	handbook	<u>11</u>	<u>12</u>	4.5	5.1	come	<u>95</u>	<u>215</u>	0.2	1.3
enroll	<u>121</u>	<u>110</u>	6.3	6.4	dean	<u>12</u>	<u>22</u>	3.9	5.1	throw	<u>9</u>	<u>13</u>	0.1	0.6
lecturer	<u>111</u>	<u>72</u>	6.2	5.8	department	<u>323</u>	<u>40</u>	4.8	1.8	X* across-i N	587	352	7.0	3.9
teach	<u>550</u>	<u>332</u>	6.0	5.3	admission	<u>73</u>	<u>14</u>	4.7	2.4	country	<u>260</u>	<u>134</u>	2.4	1.4
professor	<u>362</u>	<u>110</u>	5.9	4.2	faculty	<u>69</u>	<u>34</u>	4.2	3.2	UK	<u>25</u>	<u>12</u>	1.7	0.7
semester	<u>15</u>	<u>28</u>	3.8	5.1	enrollment	<u>16</u>	<u>8</u>	4.0	3.2	nation	<u>35</u>	<u>22</u>	1.4	0.8
tutor	<u>12</u>	<u>28</u>	3.4	5.0	reputation	<u>61</u>	<u>24</u>	3.8	2.5	United_States	<u>33</u>	<u>22</u>	1.3	0.7
enrollment	<u>26</u>	<u>30</u>	4.3	4.8	mission	<u>89</u>	<u>46</u>	3.5	2.6					
student	<u>954</u>	<u>606</u>	4.7	4.1	trustee	<u>12</u>	<u>8</u>	3.4	3.0					
mathematics	<u>48</u>	<u>9</u>	4.7	2.5	board	<u>140</u>	<u>59</u>	3.4	2.2					
freshman	<u>26</u>	<u>22</u>	4.5	4.6	president	<u>204</u>	<u>75</u>	3.4	2.0					
educate	<u>78</u>	<u>39</u>	4.6	3.7	curriculum	<u>21</u>	<u>14</u>	3.2	2.7					
lecture	<u>65</u>	<u>29</u>	4.5	3.5	budget	<u>80</u>	<u>21</u>	3.2	1.3					
faculty	<u>84</u>	<u>83</u>	4.3	4.4	commitment	<u>65</u>	<u>20</u>	3.1	1.5					
tenure	<u>31</u>	<u>10</u>	4.4	3.0	library	<u>50</u>	<u>14</u>	3.1	1.3					
instructor	<u>10</u>	<u>36</u>	2.2	4.2	policy	<u>173</u>	<u>86</u>	2.5	1.5					
undergraduate	<u>28</u>	<u>8</u>	4.2	2.6	offer	<u>25</u>	<u>8</u>	2.4	0.8					
major	<u>29</u>	<u>13</u>	4.1	3.2	program	<u>151</u>	<u>68</u>	2.2	1.1					
fellowship	<u>12</u>	<u>16</u>	3.4	4.1	chapter	<u>20</u>	<u>8</u>	2.1	0.8					
administrator	<u>41</u>	<u>21</u>	4.0	3.1	ceremony	<u>18</u>	<u>10</u>	2.0	1.2					
psychology	<u>35</u>	<u>11</u>	3.9	2.4	website	<u>29</u>	<u>17</u>	1.9	1.1					
dean	<u>8</u>	<u>13</u>	2.9	3.9	procedure	<u>29</u>	<u>14</u>	1.8	0.7					
course	<u>293</u>	<u>282</u>	3.9	3.8	plan	<u>95</u>	<u>35</u>	1.7	0.3					

13

Collocations example from the Arabic Internet Corpus

Corpus: I-AR-LEMMA; Tokens: 193842936

Query: [word="الجامعات"]

Colloc: left=0, right=1; Filter:

[LL score](#)

LL score

Collocation	Joint Freq1	Freq2	LL score	Concordance
جامع مرحل	652 77355	29458	1747.99	Examples
جامع معاهد	184 77355	9982	473.49	Examples
جامع عرب	219 77355	152191	286.37	Examples
جامع مصر	209 77355	140791	276.40	Examples
جامع حكومي	133 77355	22059	267.59	Examples
جامع مدارس	94 77355	15693	188.70	Examples
جامع السعودية	105 77355	31312	180.57	Examples
جامع أمريكي	119 77355	57162	176.73	Examples
جامع خاص	131 77355	111227	158.22	Examples
جامع الأميركية	89 77355	24380	156.75	Examples
جامع ومراكز	50 77355	1501	143.41	Examples
جامع أردن	76 77355	30729	119.21	Examples
جامع وأساتذتها	24 77355	45	105.88	Examples
جامع عراق	95 77355	102740	103.58	Examples
جامع العربية	61 77355	21642	99.58	Examples
جامع الاردنية	38 77355	2254	95.90	Examples
جامع والمؤسسات	41 77355	3827	94.14	Examples
جامع الإسلامية	62 77355	38686	84.03	Examples
جامع رسم	59 77355	47107	72.89	Examples
جامع سور	48 77355	38441	59.21	Examples
جامع بريطاني	47 77355	36262	58.82	Examples
جامع والمراكز	19 77355	737	52.01	Examples
جامع كل	113 77355	533825	48.15	Examples
جامع راسب	15 77355	342	45.11	Examples
جامع أجنب	30 77355	16253	42.70	Examples
جامع يمن	25 77355	8201	41.75	Examples

[Back to the query window](#)

See 94 examples of 'MU(meet [lemma='جامع'] [lemma='1 0- [معاهد] cut 100' in I-AR-LEMMA

- ☐ >> العليا ، الى الغاء صندوق دعم الطلاب ، وإيلاء مهامه الى **بالجامعات والمعاهد** دعا طلاب الحركة الشعبية لتحرير السودان 23 / 07 / 2008
☐ >> الأكاديمية في اليمن... و يعتبر يوم 31 مايو 2008 م الموعد **الجامعات والمعاهد** أو مقرات (www. t otal - ep - yemen. c om) على الانترنت
☐ >> استطاع لمصرأوى: الإباحية والملابس الفاضحة وراء ارتفاع - **بالجامعات والمعاهد** وزير التعليم العالي يعلن نتيجة تنسيق المرحلة الأولى للقبول
☐ >> العلمية نصيبها من هذا الاهتمام. ودعا معاليه إلى اتباع نهج **الجامعة ومعاهدها** ودعمت التعليم وأولت كافة قطاعاته الاهتمام اللازم ، وأخذت
☐ >> استطاع لمصرأوى: الإباحية والملابس الفاضحة وراء ارتفاع نسبة **بالجامعات والمعاهد** وزير التعليم العالي يعلن نتيجة تنسيق المرحلة الأولى للقبول
☐ >> العليا المتخصصة في العلوم والتقنية ومراكز البحوث العلمية **الجامعات والمعاهد** تحدد مسميات وأنشطة مؤسسات هذا التعليم. فمن هذه المؤسسات
☐ >> الجديدة [10] وعبر مدن التراب الجزائري الشاسع. وطبعاً من **الجامعات والمعاهد** ما من تكوينهم. إضافة إلى هذا نجد هؤلاء الأساتذة محترفين عبر
☐ >> استطاع لمصرأوى: الإباحية والملابس الفاضحة وراء ارتفاع - **بالجامعات والمعاهد** وزير التعليم العالي يعلن نتيجة تنسيق المرحلة الأولى للقبول
☐ >> العليا والمدارس الثانوية وحركات الشبيبة العمالية ، وحركات الشبيبة في **الجامعات والمعاهد** هذه اللجان في القرى والمدن والمخيمات ، وحركات الشبيبة في
☐ >> العالية من مواليد عام واحد من 8 % عام 1960 إلى 30 % في **الجامعات والمعاهد** وتأهيل العلماء الشباب. وهكذا ارتفعت نسبة المنتسبين إلى
☐ >> عليا متخصصة مطلباً أساسياً للتمييز في أي حقل من حقول الدراسة **جامعات ومعاهد** وفق قوله. يعتبر البحث العلمي في مؤسسات التعليم العالي من ، ؛
☐ >> استطاع لمصرأوى: الإباحية والملابس الفاضحة وراء ارتفاع نسبة **بالجامعات والمعاهد** وزير التعليم العالي يعلن نتيجة تنسيق المرحلة الأولى للقبول
☐ >> التدريب التقني ومعاهد التدريب المهني مهمة تنمية الموارد **الجامعات ومعاهد** لمراقبة المصرف المركزي ومركز الاستثمارات جميعها. 83. وتتولى
☐ >> ه الموافق 25 / 08 / 2008 م مملكة البحرين 24 / 08 / 1429 - (**الجامعات والمعاهد**) ه الموافق 08 / 10 / 2008 م إندونيسيا 08 / 10 / 1429 -
☐ >> صلاحية تحديد مواعيد إجراء الامتحانات النهائية للعام الدراسي **الجامعات والمعاهد** الامتحانات النهائية منحت وزارة التعليم العالي والبحث العلمي
☐ >> استطاع لمصرأوى: الإباحية والملابس الفاضحة وراء ارتفاع نسبة **بالجامعات والمعاهد** وزير التعليم العالي يعلن نتيجة تنسيق المرحلة الأولى للقبول
☐ >> تخرج المفكرين والعلماء ، وتضرب مثلاً رائعاً في التصدي **و جامعات ومعاهد** استطاعوا هزيمة السجان وأن يحولوا السجون والزنازين إلى مدارس
☐ >> فحسب ، ولكنه قصد أيضاً الإشارة إلى بيت الأسرة باعتباره **والجامعات والمعاهد** فالكتاب لم يقصد بمصطلح المؤسسات التعليمية الإشارة للمدارس .
☐ >> المتخصصة ، على أن يصحب ذلك برنامج للسعودة ، من حيث نوع **الجامعات والمعاهد** في مجال عملها ، ودعم البرامج التدريبية في المجال ذاته ، في
☐ >> التعليم العالي العراقية لعب دور ريادي في المجتمع العراقي **الجامعات ومعاهد** أسس السلام والاستقرار على المدى البعيد. ويقول إن بمقدور
☐ >> الحكومية للعام الدراسي المقبل. وبعد اكتشاف الخطأ قامت **الجامعات والمعاهد** مفاضلة أعضاء الهيئة التدريسية وأبناء الشهداء من أجل دخول
☐ >> (هو منهج هذه البلاد القائم على الوسطية على حد قوله تعالى **الجامعة ومعاهدها** إلى الاستعداد لاستقبالها ، ومتبيراً إلى أن المنهج المتبع في
☐ >> التابعة للمجلس الأعلى للمعاهد لتبدأ اعتباراً من 6 / 6 / 2007 **الجامعات والمعاهد** اليوم السبت قراراً بتأجيل امتحانات الفصل الدراسي الثاني في
☐ >> 14. كان هناك مجموعتان من الأطفال يلعبون بالقرب من **والجامعات والمعاهد** القومي للاتصالات إلى جانب شركات المحمول والقطاع الخاص
☐ >> العليا مركز القاهرة لدراسات حقوق الإنسان - 17 / 7 / 2005 **الجامعات والمعاهد** الدورة التدريبية السنوية الثانية عشر لحقوق الإنسان لطلاب

Oxford bilingual dictionaries

Translate

English to French

SEARCH BROWSE

university GO

help

more search options

Found **university** in 50 entries

- university**
... degree town universitaire place à l'université **university** entrance entrée f à l'université university ...
- Open University**
... OU Open **University** n British English University enseignement universitaire par correspondance ...
- PPE**
... British English **University** philosophy, politics and economics philosophie, politique et économie f ...
- redbrick university**
n British English université f autre que Cambridge et Oxford ...
- free university**
n American English University système d'enseignement universitaire par correspondance ouvert à tous ...

Results 1 to 5 of 50

pages/4/2011

1 2 3 4 5 6 7 8 9 10 next >

outline view

show/hide options...

☒ Examples

☒ Idioms

☒ Phrasal Verbs

print | email | cite | link

French and English and resources

back to results

university [juːnɪˈvɜːsəti] Br Am English

A. n **POS**

université (f) French

B. modif

[lecturer] **d'université**; [degree / town] **universitaire**; [place] **à l'université**

university **entrance** = entrée (f) à l'université

university **education** = formation (f) universitaire

More on this word

university in other languages:

German	Universität (f)
Italian	università (f)
Portuguese	universidade (f)
Spanish	universidad (f)
Japanese	大学
Korean	(종합) 대학교
Polish	uniwersytet (m)
Russian	университет (m)
Chinese	综合性大学
Thai	มหาวิทยาลัย
Arabic	جامعة

See Language Web display

Find this word in Oxford Reference Online

Examples from parallel corpora

Other languages

Google translation

Google translate

From: Arabic To: English Translate

جامعة

☒ Allow phonetic typing

Arabic to English translation

University

Dictionary - [View detailed dictionary](#)

noun

1. university
2. academy
3. varsity

abbreviation

1. uni
2. u-ty

Dictionary

Found in dictionary: Arabic > English.

☆ جامعة

noun

1. academy
2. varsity
3. university

abbreviation

1. uni
2. u-ty

Related phrases

[مدير جامعة](#)

chancellor

[طرد من جامعة](#)

send down

[رئيس جامعة مستشار أو عميد](#)

chancellor

[عضو في إدارة جامعة بريطانية](#)

fellow

[طالب في جامعة](#)

commoner

[خريج جامعة](#)

fellow

[بيان جامعة](#)

calender

[تخرج من جامعة](#)

graduate

[مكتب رئاسة جامعة](#)

rectorship

[رئاسة جامعة](#)

rectorship

Oxford bilingual dictionaries

DICTIONARIES **OXFORD LANGUAGE WEB**

Translate
 Arabic

جامعة **GO**

help

Word list

more ▲

جَامِعَة noun

جَانِب noun

جَانِب الطَّرِيق noun

جَاهِز adjective

جَاهِزَة adjective

جَبَابِر noun

جِبَال noun

جَبَل noun

جِبْنَة noun

جَنْهَة noun

جَبِيرَة noun

جَبِين noun

جَنَّة noun

جُنُث noun

جُحْر noun

جُحُور noun

جَد noun

جَدَّ noun

جَدَّاد noun

Dictionary entry

جَامِعَة

► institution which provides a high level of education for sb who has left school

Meaning in English

Language

German: Universität *noun*

Italian: università *noun*

Polish: uniwersytet *noun*

Russian: университет *noun*

Thai: มหาวิทยาลัย mahawiththayalai

Chinese: 综合性大学 zōnghéxing dàxué

Korean: (종합) 대학교 (jonghap) daehakgyo

Japanese: 大学 daigaku

Arabic: جَامِعَة *noun*

POS

Pronunciation

Dictionary entry

جامعة *noun*

Oxford bilingual dictionaries

- Do users need to know the meaning of word in many different languages?
- Users need to translate from one language to another.
- Connecting terms with a central language (English) can connect two languages together.
- specific linguistic information to each word from their monolingual dictionaries can be provided, in addition to information of the central language (English)

Arabic & Arabic NLP

- 200 million people speaking Arabic as first language.
- More than 1 billion Muslims need Arabic to recite the Quran (the holy book of Muslims).
- One of the UN official languages.
- Increased potential for learning Arabic recently.
- ***Many commercial software companies invest in Arabic NLP.***

Why is Arabic NLP difficult?

- **Complex morphology**

- Words consist of multi-morphemes of 5 kinds
- Proclitics, prefixes, stem/root, suffixes, enclitics.

- وَسَيَكْتُبُوهَا [*wasayaktubūnhā*] (*And they will write it*)

و	سَ	يَ	كُتِبَ	وَنَ	هَآ
Conjunction	Particle of futurity	Progressive letter	Root / Stem	Relative Pronoun (Plural/Subject)	Relative Pronoun (Object)

- **Vowels & Diacritic Marks**

- 3 long vowels (ا alif, و wāw, ي yā')
- 3 short vowels (fatha^h َ , damma^h ُ , kasra^h ِ)
- Other diacritics: sukūn ْ , šadda^h ّ , tanwīn (ً , ٌ , ٍ)
- taṭwīl character (-).
- hamza^h (ء , ا , إ , ئ), tā' marbūṭa^h (ة) and hā' (ه), yā' (ي) and alif maqṣūra^h (ى), and madda^h (ّ).

Morphological Analyses of Arabic text

- Morphological analysis is essential for processing text corpora and building dictionaries.
- Existing Arabic morphological analyzers failed to achieve accuracy rates more than 75%. (Sawalha & Atwell, 2008)
- We can not rely on such analyzers for further analysis such as part-of-speech tagging and parsing.

Morphological Analyzer for Arabic text

Step 1: Tokenization

- Different text types
- Spell-checking

Step 2: Function words

Step 3: Clitics, Affixes & Stems

Step 4: Root/Lemma extraction

Step 5: Pattern generation

Step 6: Vowelization

Step 7: Assigning detailed morphological features tags for each of the word's morphemes

Example of Analyzed word

وَالْجَامِعَاتُ *wa al-ġāmi‘āt* (And the universities)

ا ت	جَامِعَ	الْ	وَ
Feminine plural letters	stem	Definite article	conjunction
Lemma <link> <u>جَامِعَة</u>	Pronunciation Lemma root POS Pattern Meaning Examples Phrases, collocations, idioms Origin (links to traditional Arabic lexicons)	Root <link> <u>جَمَعَ</u>	Plural form <u>جَامِعَات</u>
Pattern <u>فَاعِلَة</u>	Word's morphemes		
وَ	p--c-----	Particle, conjunction (clitic)	
الْ	r---d-----	definite article (clitic)	
جَامِعَ	np----flp-vn-dd---ncat-s	collective noun, feminine (M/S), varied, non-human ...	
ا ت	r---l-----	plural feminine letters	

12/4/2011

Word's list of similar root

الْجَمْعُ
 التَّجْمِيعُ
 اجْتِمَاعُ
 إِجْمَاعُ
 جَمَعَ
 تَجَمُّعُ
 جَامِعُ
 جَامِعِي
 جَامِعِيُونَ
 جَامِعِيَّةُ
 جَامِعِيَّاتُ
 جَمْعِيَّةُ
 مَجْمَعُ
 مَجْمَعُ
 مَجْمُوعُ
 يَجْمَعُ

Samples of traditional Arabic lexicons

كُتِبَ: الكتابُ: معروف، والجمع كُتُبٌ وكُتُبٌ. كُتِبَ الشيءَ يَكْتُبُهُ كُتُبًا وكِتَابَةً، وكُتِبَ: خَطَّه؛ قال أبو النجم: أَقْبَلْتُ مِنْ عِنْدِ زِيَادٍ كَاخِرَفٌ، تَخَطَّ رَجُلَايَ بِحُطٍّ مُخْتَلِفٍ، تُكْتَبَانِ فِي الطَّرِيقِ لَامَ أَلِفٍ قَالَ: وَرَأَيْتُ فِي بَعْضِ النُّسخِ تُكْتَبَانِ، بِكسر التاء، وهي لغة بهراء، يَكْسِرُونَ التاءَ، فيقولون: تَعْلَمُونَ، ثُمَّ أَتْبَعَ الكافَ كسرةَ التاءِ. وَالْكِتَابُ أَيْضًا: الاسمُ، عن اللحياني. الأزهري: الكتابُ اسم لما كُتِبَ مَجْمُوعًا؛ وَالْكِتَابَةُ لِمَنْ تَكُونُ لَهُ صِنَاعَةٌ، مثل الصَّيَاغَةِ وَالْحَيَاظَةِ. وَالْكُتْبَةُ: اِكْتَابُكَ كِتَابًا تنسخه. ويقال: اِكْتَبَ فلانُ فلانًا أَي سألَه أَنْ يَكْتُبَ لَهُ كِتَابًا فِي حاجة. وَاسْتَكْتَبَهُ الشيءَ أَي سألَه أَنْ يَكْتُبَهُ لَهُ. ابن سيده: اِكْتَبَهُ كُتِبَ. وقيل: كُتِبَ خَطَّه؛ وَاِكْتَبَهُ: اسْتَمْلَاهُ، وكذلك اسْتَكْتَبَهُ. وَاِكْتَبَهُ: كُتِبَ، وَاِكْتَبَهُ: كُتِبَ. وفي التزويل العزيز: اِكْتَبَهَا فهي تُمْلَى عَلَيْهِ بُكْرَةً وَأَصِيلًا؛ أَي اسْتَكْتَبَهَا. ويقال: اِكْتَبَ الرجلُ إِذَا كُتِبَ نَفْسَهُ فِي دِيوانِ السُّلْطَانِ ...

k t b: [*Alkitab*] the book; is well known. The plural forms are [*kutubun*] and [*kutbun*]. [*kataba Alshay'*] He wrote something. [*yaktubuhu*] the action of writing something. [*katban*], [*kitabān*] and [*kitabatan*] means the art of writing. And [*kattabahu*] writing it means draw it up. Abu Al-Najim said: I returned back from Ziyad's house [after meeting him] and behaved demented, my legs drawn up differently (means walking in a different way). They wrote [*tukattibani*] on the road the letters of *Lam Alif* (describing how he was walking crazily and in a different way). He said: I saw in a different version, the word "they wrote" [*tikittibani*] using the short vowel *kasrah* on the first letter [taa], as it is used by Bahraa' (Arab tribe) dialect. They say: (ti'lamuwn) (you know). Then the short vowel *kasrah* is propagated to the following letter (kaf). Moreover, [*Alkitab*] the book is a noun. Al-lihyani Al-Azhari definition is: [*Alkitab*] The book is the name of a collection of what has been written (a collection of written materials or texts). And the book has gerund [*Alkitabatu*] writing (art of writing) for whoever has a profession, similar to drafting and sewing. And [*Alkitabatu*]: is copying a book [copying a book in several copies]. It is said: [*iktataba*] someone subscribed another means; he asked to write him a letter in something. [*istaktabahu*] He dictated someone something means to write him something. Ibn Sayyedah: [*iktatabahu*] is similar to [*katabahu*]. It is said: [*katabahu*] write something down means draw up. And [*iktatabahu*] writing something down means dictate someone something, which is the same meaning of [*Istaktabahu*]. [*Iktatabahu*] registering (masculine), and [*Iktatabathu*] registing (feminine). In the Qur'an: [*Iktatabaha*] He registered it, he has dictated it every sunrise and sunset, which means dictating it. It is said: [*Iktataba Al-rajul*] The man registered, if he registered himself in the Sultan's office ...

Constructing the Arabic broad-coverage lexical resource

Analyzing Lexicon's Text Separately

- Convert each lexicon text into a unified format.

*** كُتِبَ: الكتابُ معروف، والجمع كُتُبٌ وكُتُبٌ. كُتِبَ الشيءَ يَكْتُبُهُ كِتَابًا وكِتَابَةً، وكُتِبَ: خَطَّهُ؛ قال أبو النجم: أَقْبَلْتُ من عِنْدِ زِيَادٍ كَالْحَرْفِ، تَخْطُ رَجُلَايَ بِحُطِّ مُخْتَلِفٍ، تُكْتَبَانِ في الطَّرِيقِ لَامَ أَلْفٍ قال: ورأيت في بعض النسخ تُكْتَبَانِ بكسر التاء، وهي لغة بهراء، يَكْسِرُونَ التاءَ، فيقولون: تَعْلَمُونَ، ثم أَتَعَ الكافَ كَسْرَةَ التاء. والكتابُ أيضًا: الاسمُ، عن اللحياني. الأزهري: الكتابُ اسم لما كُتِبَ مجموعاً؛ والكتابُ مصدر؛ والكتابةُ لِمَنْ تَكُونُ لَهُ صِنَاعَةٌ، مثل الصَّيَاغَةِ وَالْحِصَاظَةِ. والكُتَيْبَةُ: اِكْتِابُكَ كِتَابًا تنسخه. ويقال: اِكْتَبَ فلان فلاناً أي سأله أن يَكْتُبَ له كِتَابًا في حاجة. واِسْتَكْتَبَ الشيءَ أي سأله أن يَكْتُبَ له. ابن سيده: اِكْتَبَهُ كُتَيْبَةً. وقيل: كُتِبَ خَطَّهُ؛ واِكْتَبَهُ: اِسْتَمْلَاهُ، وكذلك اِسْتَكْتَبَهُ. واِكْتَبَهُ: كُتِبَ، واِكْتَبَهُ: كُتِبَ. وفي التزويل العزيز: اِكْتَبَهَا فهي تُمْلَى عليه بكرةً وأصيلاً؛ أي اِسْتَكْتَبَهَا. ويقال: اِكْتَبَ الرجل إذا كُتِبَ نفسه في ديوانِ السُّلْطَانِ ...

- A bag of words is extracted from the definition text.

(الكتابُ ، كتب)	(كُتِبَ ، كتب)	(وكتابةً ، كتب)	(النجم ، كتب)	(كالحرف ، كتب)	(تُكْتَبَانِ ، كتب)	(قال ، كتب)
(معروف ، كتب)	(الشيء ، كتب)	(وكتبه ، كتب)	(أَقْبَلْتُ ، كتب)	(تَخْطُ ، كتب)	(في ، كتب)	(ورأيت ، كتب)
(والجمع ، كتب)	(يَكْتُبُهُ ، كتب)	(خَطَّهُ ، كتب)	(من ، كتب)	(رَجُلَايَ ، كتب)	(الطَّرِيقِ ، كتب)	(في ، كتب)
(كُتِبُ ، كتب)	(كُتِبًا ، كتب)	(قال ، كتب)	(عِنْدِ ، كتب)	(بِحُطِّ ، كتب)	(لَامَ ، كتب)	(بعض ، كتب)
(كُتِبُ ، كتب)	(وكتاباً ، كتب)	(أبو ، كتب)	(زيادٍ ، كتب)	(مُخْتَلِفٍ ، كتب)	(أَلْفٍ ، كتب)	(النسخ ، كتب)

- A normalization analysis that verifies the word-root pairs by applying linguistic knowledge.

أَكْتَبَهُ	'aktabahu	الكتاب	al-kitāb	الْكُتْبَةُ	al-kutba ^{tu}
أَكْتُبَ	'aktaba	الكتابة	al-kitāba ^t	الْكُتْبَةُ	al-kutba ^{tu}
أَكْتُبْتُ	'aktabtu	الكتابة	al-kitāba ^{ta}	الكتاب	al-kitāb
أَكْتُبْنِي	'aktibnī	الكتابة	al-kitāba ^t	الكتابة	al-kitāba ^{tu}
إِكْتَابًا	'iktāb ^{an}	الكتائب	al-katātīb	الكتاب	al-kitāba
اسْتُكْتُبَهُ	'istaktabahu	الكتبة	al-kitba ^t	الكتابة	al-kitāba ^{tu}
اسْتُكْتُبَهُ	'istaktabahu	الكتيبة	al-katība ^t	الكتاب	al-kitābu
اسْتُكْتُبَهَا	'istaktabahā	وكتيبة	wa katība ^t	الكتاب	al-kitābi
اِكْتُبْ	'iktataba	الكتائب	al-katā'iba	المكاتب	al-mukātib
اِكْتُبْ	'iktataba	الكتائب	al-katā'ibu	المكاتبة	al-mukātiba ^t
اِكْتُبَهُ	'iktatabahu	الكتيبة	al-katība ^{ta}	المكتب	al-maktab
اِكْتُبَهَا	'iktatabahā	الكتائب	al-katā'iba	المكتبة	al-maktaba ^t
اِكْتُبْ	'uktub	الكتبة	al-kataba ^t	المكتوبة	al-maktūba ^t
اِكْتُبْتُ	'uktutibtu	الكتب	al-katbu	الْكُتُبُ	al-kuttābu
اِكْتُبْكَ	'iktitābuk	الكتب	al-katbi	الْكِتَابُ	al-kitāba
اِكْتُبْكَ	'iktitābuka	الْكُتُبُ	al-kutabu	الْكِتَابَةُ	al-kitāba ^{tu}
الْاِكْتُتَابُ	al-'iktitābu	الْكُتَيْبَةُ	al-kutaybatu	الْكِتَابَةُ	al-kitāba ^{ti}
التكاتب	at-takātubu	الْكُتُبُ	al-kuttāba	الْمَكْتُبُ	al-maktabu
الكاتب	al-kātib	الْكُتُبُ	al-kuttābi	الْمَكْتُوبَةُ	al-maktūba ^{tu}
الكاتب	al-kātibu	الْكُتْبَةُ	al-kutba ^t	اِسْتُكْتُبْ	'istaktaba

Combining into One Broad-Coverage Lexical Resource

- After analyzing each lexicon, a combination algorithm is applied to construct the broad-coverage lexicon.

#	Lexicon	Word types [B]	Records inserted [A]	Percentage (A/B)% (A/C)%	
1	<i>lisān al-‘rab</i>	207,992	207,992	100.00%	47.80%
2	<i>mu’ğam al-muḥīṭ fī al-luḡat</i>	74,507	61,113	82.02%	14.04%
3	<i>tağ al-‘arūs min ḡawāhir al-qāmūs</i>	128,119	95,415	74.47%	21.93%
4	<i>muḥīṭ aṣ-ṣiḥāḥ</i>	19,540	16,573	84.82%	3.81%
5	<i>al-muḡrab fī tartīb al-mu‘rab</i>	12,396	9,805	79.10%	2.25%
6	<i>kitābu al-‘ayn</i>	30,292	18,878	62.32%	4.34%
7	<i>al-mu’ğam al-wasīṭ</i>	36,660	25,364	69.19%	5.83%
Totals		509,506	435,140 [C]	85.40%	100.00%

The Corpus of Traditional Arabic Lexicons

- Lexicons' text can be used as a **corpus of traditional Arabic lexicons**.
- **Different domain** than existing corpora.
- The Arabic corpus of dictionaries **covers a period of more than 1200 years**.
- Consists of **large number of words and word types**.
- Has both **vowelized and non-vowelized text**.

Number of files		247
Size		178.32 MB
Vowelized words analysis	# of words	14,369,570
	# of word types	2,184,315
Non-vowelized word analysis	# of words	14,369,570
	# of word types	569,412

Partially-vowelized		Non-vowelized	
Word	Frequency	Word	Frequency
في	292,396	من	322,239
من	269,200	في	301,895
قال	172,631	قال	190,918
و	120,060	أي	132,635
على	108,252	و	130,809
ما	89,195	على	119,639
وقال	88,233	إذا	115,842
عن	82,027	وقال	99,601
إذا	81,479	ابن	94,980
أي	78,622	ما	94,530
وهو	75,149	بن	92,213 ²⁷

Key notes

- When do users start using dictionaries?
- What are the users' needs of monolingual or bilingual dictionaries?
- What is the best arrangement methodology for Arabic dictionaries?
- What are the technologies needed to search for a word in dictionary?
- What linguistic information that can be provided in a dictionary for each dictionary entry?
- How is dictionary entry connected to other dictionary entries?
- How are examples of the dictionary entry selected?
- What are the phrases, collocations, idioms that are related to dictionary entry?

Conclusions

- Morphological analysis is essential technology for constructing dictionaries.
- The Arabic morphological analyzer provides detailed information about the processed words.
- Traditional Arabic lexicons are important because they provide the correct linguistic information about the words, examples from the Qur'an and poetry, collocation used in the past 1200 years, etc...
- Arrangement of Arabic dictionary entries is a challenge for Arabic lexicography.
- Three types of dictionaries can be provided according the user needs; Modern Arabic dictionary, traditional Arabic dictionary, and a dictionary of dialects.

References

- Sawalha, MS; Atwell, ES *Constructing and Using Broad-coverage Lexical Resource for Enhancing Morphological Analysis of Arabic* in: **Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)**, pp.282-287. European Language Resources Association (ELRA). 2010.
- Sawalha, M; Atwell, ES *Fine-Grain Morphological Analyzer and Part-of-Speech Tagger for Arabic Text* in: **Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)**, pp.1258-1265. European Language Resources Association (ELRA). 2010.
- Sawalha, M; Atwell, ES (*توظيف قواعد النحو والصرف في بناء محلل صرفي للغة العربية*) *Adapting Language Grammar Rules for Building a Morphological Analyzer for Arabic Text* in: **Proceedings of ALECSO Arab League Educational Cultural and Scientific Organization workshop on Arabic morphological analysis**. 2009.
- Sawalha, M; Atwell, ES *Linguistically Informed and Corpus Informed Morphological Analysis of Arabic* in: **Proceedings of CL2009 International Conference on Corpus Linguistics**. 2009.
- Sawalha, M; Atwell, E *Comparative evaluation of Arabic language morphological analysers and stemmers* in: **Proceedings of COLING 2008 22nd International Conference on Computational Linguistics (Poster Volume)**, pp.107-110. 2008.

Thank you