CS-AD 220 – Spring 2016

Natural Language Processing

Session 1: 28-Jan-16

Prof. Nizar Habash

Some slides are adapted from Jurafsky and Martin's course slides on Speech and Language Processing

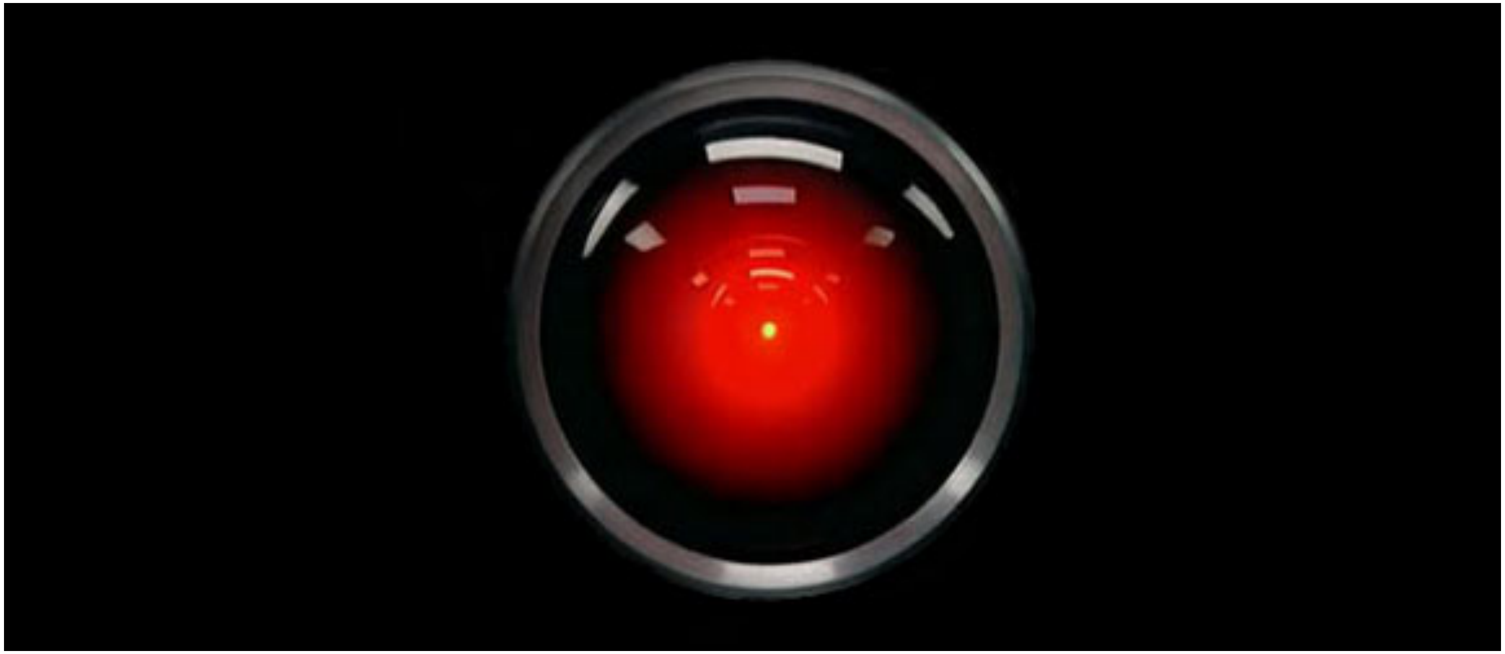
Roadmap

- Introduction
  - **Professor & Students**
  - Topics
  - Syllabus discussion
  - Books and Resources
- Lecture 1

Roadmap

- Introduction
  - Professor & Students
  - **Topics**
  - Syllabus discussion
  - Books and Resources
- Lecture 1

Sci-Fi: HAL 9000



- https://www.youtube.com/watch?v=HwBmPiOmEGQ
- Discussion:
  - What are the challenges in developing a system like HAL 9000?

Natural Language Processing

- Also known as
  - Computational Linguistics
  - Human Language Technology
- NLP is an interdisciplinary field
  - Computer science
  - Linguistics
  - Cognitive science, psychology, pedagogy, mathematics, etc.
- Applied natural language processing
  - Develop practical applications modeling human languages
- Theoretical computational linguistics
  - Focus on theoretical linguistics and cognitive science

Natural Language Processing

- Applications
  - Machine Translation (MT)
  - Information Retrieval (IR)
  - Automatic Speech Recognition (ASR)
  - Optical Character Recognition (OCR)
  - Automatic Summarization, Speech Synthesis, etc.
- Enabling Technologies
  - Tokenization
  - Part-of-Speech Tagging
  - Syntactic Parsing
  - Lemmatization
  - Word Sense Disambiguation, etc.

Natural Language Processing

- Rule-based/Symbolic Approaches
  - Linguists write rules that are applied by the machines
- Corpus-based/Statistical Approaches
  - Machines learn the "rules" from training data
    - Annotated data – supervised methods
      - Parallel Corpora: translated text collections
      - Treebanks: manually syntactically analyzed texts
      - Speech Corpora with transcripts
    - Unannotated data – unsupervised methods
    - Semi-supervised methods
  - Machine learning approaches are dominant in the field
- Hybrid Approaches
  - The best of **Smart**/Slow Humans and Dumb/**Fast** Machines

Class Topics

- Basic text processing
- Finite state machines and word morphology modeling
- N-gram language modeling
- Part-of-speech tagging
- Syntactic parsing
- Machine translation
- A review of sentiment analysis, information retrieval, and question answering

Assignments

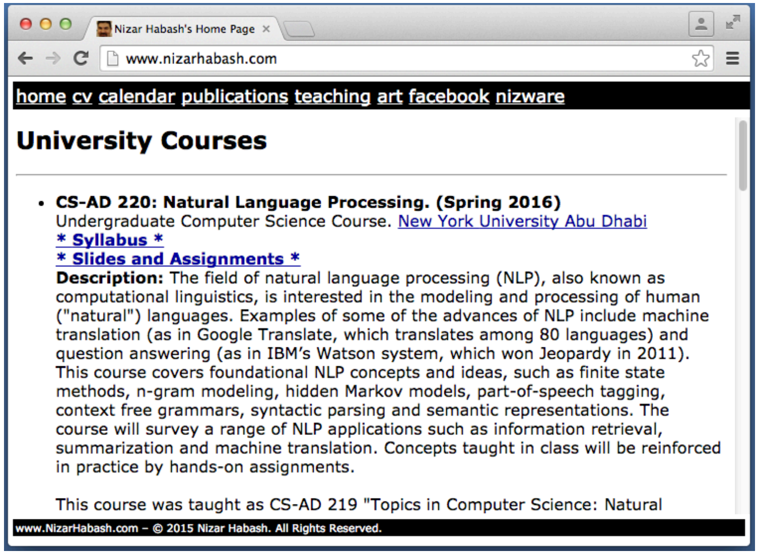
1. Basic text processing
2. Finite state machines
3. Part-of-speech tagging and parsing
4. Machine translation

Roadmap

- Introduction
  - – Professor & Students
  - – Topics
  - – **Syllabus discussion**
  - – Books and Resources
- Lecture 1

Syllabus

- Online:
  - https://sites.google.com/a/nyu.edu/nyuad-cs-ad-220-natural-language-processing-spring-2016/
  - Go to my website: http://www.nizarhabash.com/
    - Then click on **teaching**

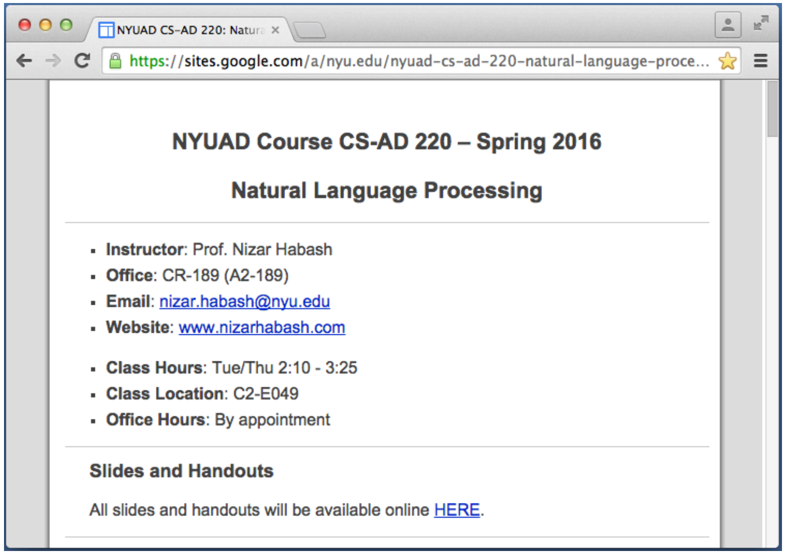


University Courses

- **CS-AD 220: Natural Language Processing. (Spring 2016)**
  Undergraduate Computer Science Course. New York University Abu Dhabi
  *** Syllabus ***
  *** Slides and Assignments ***
  Description: The field of natural language processing (NLP), also known as computational linguistics, is interested in the modeling and processing of human ("natural") languages. Examples of some of the advances of NLP include machine translation (as in Google Translate, which translates among 80 languages) and question answering (as in IBM's Watson system, which won Jeopardy in 2011). This course covers foundational NLP concepts and ideas, such as finite state methods, n-gram modeling, hidden Markov models, part-of-speech tagging, context free grammars, syntactic parsing and semantic representations. The course will survey a range of NLP applications such as information retrieval, summarization and machine translation. Concepts taught in class will be reinforced in practice by hands-on assignments.

  This course was taught as CS-AD 219 "Topics in Computer Science: Natural

www.NizarHabash.com – © 2015 Nizar Habash. All Rights Reserved.


NYUAD CS-AD 220: Natur
https://sites.google.com/a/nyu.edu/nyuad-cs-ad-220-natural-language-proce...

NYUAD Course CS-AD 220 – Spring 2016

Natural Language Processing

- **Instructor**: Prof. Nizar Habash
- **Office**: CR-189 (A2-189)
- **Email**: nizar.habash@nyu.edu
- **Website**: www.nizarhabash.com

- **Class Hours**: Tue/Thu 2:10 - 3:25
- **Class Location**: C2-E049
- **Office Hours**: By appointment

Slides and Handouts

All slides and handouts will be available online HERE.

Roadmap

- Introduction
  - Teacher & students
  - Topics
  - Syllabus discussion
  - **Books and Resources**
- Lecture 1

Books and Resources

- **JM:** Daniel Jurafsky and James H. Martin. "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition." (2nd Edition). Pearson Prentice Hall, 2008.
  The book is available at the bookstore.

- **NH:** Nizar Habash. "Introduction to Arabic Natural Language Processing." Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers, 2010.
  This book is available online.
  URL:
  http://www.morganclaypool.com/doi/abs/10.2200/
  S00277ED1V01Y201008HLT010

Roadmap

- Introduction
  - Teacher & students
  - Topics
  - Syllabus discussion
  - Books and Resources
- **Lecture 1**

Natural Language Processing

- We're going to study what goes into getting computers to perform useful and interesting tasks involving human languages.
- We are also concerned with the insights that such computational work gives us into human processing of language.

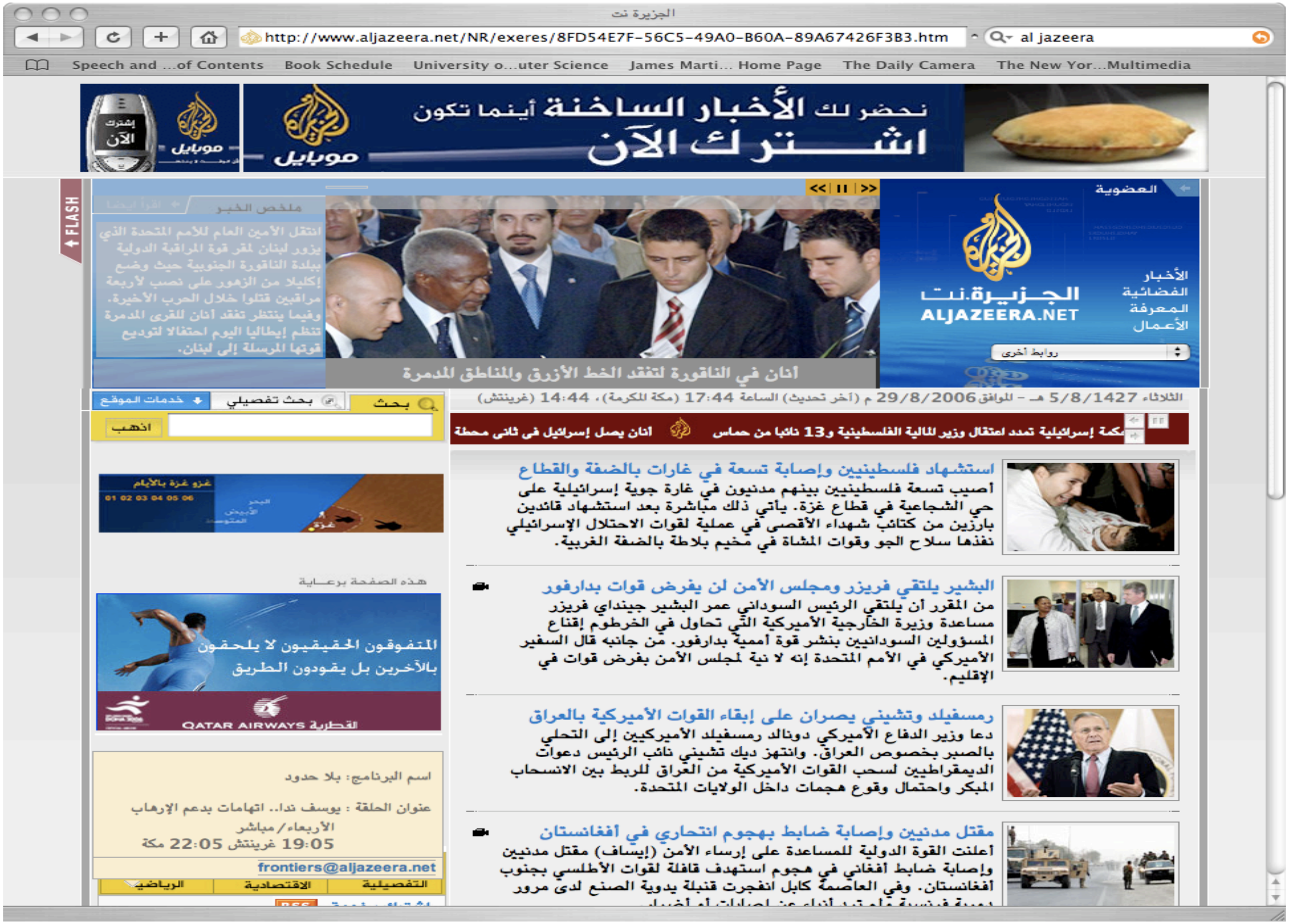
Why Should You Care?

Important trends

1. An enormous amount of knowledge is now available in machine readable form as natural language text
2. Conversational agents are becoming an important form of human-computer communication
3. Much of human-human communication is now mediated by computers

Very cool stuff! And with lots of commercial interest.

Google Translate



انتقل الأمين العام للأمم المتحدة الذي يزور لبنان لمقر قوة المراقبة الدولية ببلدة الناقورة الجنوبية حيث وضع إكليلا من الزهور على نصب لأربعة مراقبين قتلوا خلال الحرب الأخيرة. وفيما ينتظر تفقد أنان للقرى المدمرة تنظم إيطاليا اليوم احتفالا لتوديع قوتها المرسلة إلى لبنان.

أنان في الناقورة لتفقد الخط الأزرق والمناطق المدمرة

الثلاثاء 5/8/1427 هـ - الموافق 29/8/2006 م (آخر تحديث) الساعة 17:44 (مكة المكرمة)، 14:44 (غرينتش)

كمة إسرائيلية تمدد اعتقال وزير المالية الفلسطينية و13 نائبا من حماس

أنان يصل إسرائيل فى ثانى محطة

استشهاد فلسطينيين وإصابة تسعة في غارات بالضفة والقطاع

أصيب تسعة فلسطينيين بينهم مدنيون في غارة جوية إسرائيلية على حي الشجاعية في قطاع غزة. يأتي ذلك مباشرة بعد استشهاد قائدين بارزين من كتائب شهداء الأقصى في عملية لقوات الاحتلال الإسرائيلي نفذها سلاح الجو وقوات المشاة في مخيم بلاطة بالضفة الغربية.

البشير يلتقي فريزر ومجلس الأمن لن يفرض قوات بدارفور

من المقرر أن يلتقي الرئيس السوداني عمر البشير جينداي فريزر مساعدة وزيرة الخارجية الأميركية التي تحاول في الخرطوم إقناع المسؤولين السودانيين بنشر قوة أممية بدارفور. من جانبه قال السفير الأميركي في الأمم المتحدة إنه لا نية لمجلس الأمن بفرض قوات في الإقليم.

رمسفيلد وتشيني يصران على إبقاء القوات الأميركية بالعراق

دعا وزير الدفاع الأميركي دونالد رمسفيلد الأميركيين إلى التحلي بالصبر بخصوص العراق. وانتهز ديك تشيني نائب الرئيس دعوات الديمقراطيين لسحب القوات الأميركية من العراق للربط بين الانسحاب المبكر واحتمال وقوع هجمات داخل الولايات المتحدة.

مقتل مدنيين وإصابة ضابط بهجوم انتحاري في أفغانستان

أعلنت القوة الدولية للمساعدة على إرساء الأمن (إيساف) مقتل مدنيين وإصابة ضابط أفغاني في هجوم استهدف قافلة لقوات الأطلسي بجنوب أفغانستان. وفي العاصمة كابل انفجرت قنبلة يدوية الصنع لدى مرور دورية فرنسية ولم ترد أنباء عن إصابات أو أضرار.

اسم البرنامج: بلا حدود

عنوان الحلقة : يوسف ندا.. اتهامات بدعم الإرهاب

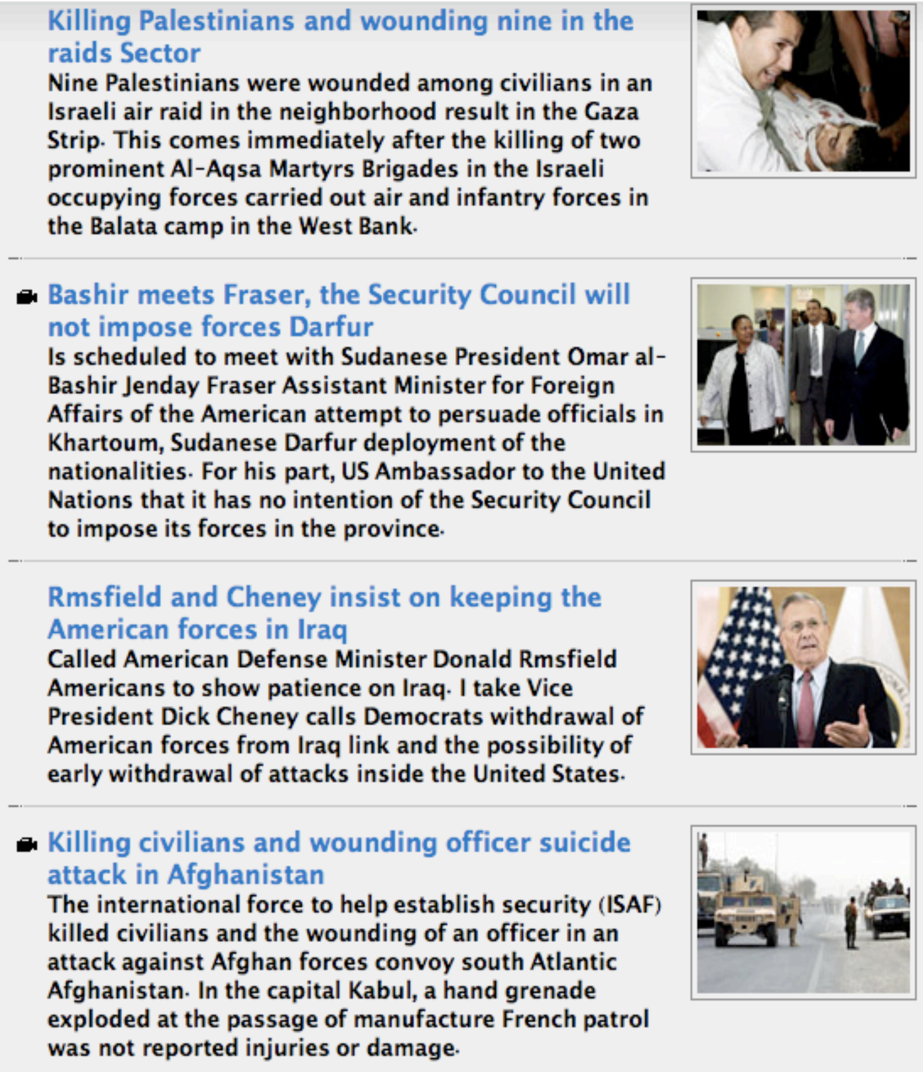
الأربعاء/مباشر

19:05 غرينتش 22:05 مكة

frontiers@aljazeera.net

Adapted from Speech and Language Processing - Jurafsky and Martin

Google Translate



Killing Palestinians and wounding nine in the raids Sector

Nine Palestinians were wounded among civilians in an Israeli air raid in the neighborhood result in the Gaza Strip. This comes immediately after the killing of two prominent Al-Aqsa Martyrs Brigades in the Israeli occupying forces carried out air and infantry forces in the Balata camp in the West Bank.

Bashir meets Fraser, the Security Council will not impose forces Darfur

Is scheduled to meet with Sudanese President Omar al-Bashir Jenday Fraser Assistant Minister for Foreign Affairs of the American attempt to persuade officials in Khartoum, Sudanese Darfur deployment of the nationalities. For his part, US Ambassador to the United Nations that it has no intention of the Security Council to impose its forces in the province.

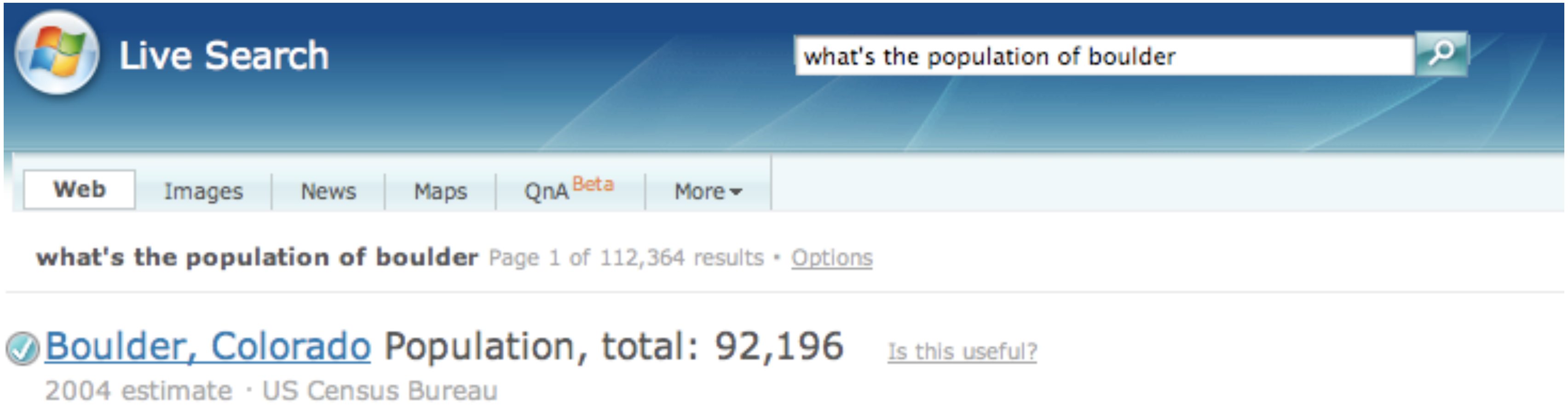
Rmsfield and Cheney insist on keeping the American forces in Iraq

Called American Defense Minister Donald Rmsfield Americans to show patience on Iraq. I take Vice President Dick Cheney calls Democrats withdrawal of American forces from Iraq link and the possibility of early withdrawal of attacks inside the United States.

Killing civilians and wounding officer suicide attack in Afghanistan

The international force to help establish security (ISAF) killed civilians and the wounding of an officer in an attack against Afghan forces convoy south Atlantic Afghanistan. In the capital Kabul, a hand grenade exploded at the passage of manufacture French patrol was not reported injuries or damage.

Web Q/A



what's the population of boulder

what's the population of boulder Page 1 of 112,364 results · Options

Boulder, Colorado Population, total: 92,196 Is this useful?

2004 estimate · US Census Bureau

Weblog Analytics

- Data-mining of Weblogs, discussion forums, message boards, user groups, and other forms of user generated media
  - Product marketing information
  - Political opinion tracking
  - Social network analysis
  - Buzz analysis (what's hot, what topics are people talking about right now).

Information Extraction & Sentiment Analysis

Attributes:

zoom

affordability

size and weight

flash

ease of use

Size and weight

- ✓ nice and compact to carry!
- ✓ since the camera is small and [illegible] eed to carry around those heavy, [illegible] nal cameras either!
- ✗ the camera feels flimsy, is plastic and very light in weight you have to be very delicate in the handling of this camera

Adapted from Speech and Language Processing - Jurafsky and Martin

Major Topics

1. Words
2. Syntax
3. Meaning
4. Discourse

5. Applications exploiting each

Adapted from Speech and Language Processing - Jurafsky and Martin

Applications

- First, what makes an application a *language processing application* (as opposed to any other piece of software)?
  - An application that requires the use of knowledge about human languages
    - Example: Is Unix wc (word count) an example of a language processing application?

Applications

- Word count?
  - When it counts words: Yes
    - To count words you need to know what a word is. That' s knowledge of language.
  - When it counts lines and bytes: No
    - Lines and bytes are computer artifacts, not linguistic entities

Big Applications

- Question answering
- Conversational agents
- Summarization
- Machine translation

Big Applications

- These kinds of applications require a tremendous amount of knowledge of language.
- Consider the interaction with HAL the computer from 2001: A Space Odyssey
  - – Dave: *Open the pod bay doors, Hal.*
  - – HAL: *I'm sorry Dave, I'm afraid I can't do that.*