Information Extraction

CIS, LMU München
Winter Semester 2016-2017

Dr. Alexander Fraser, CIS
Information Extraction – Administravia - I

- Vorlesung
  - Learn the basics of Information Extraction (IE)

- Seminar
  - Each student will present a Referat on IE (Powerpoint, LaTeX, Mac)
    - The group will discuss it
  - Also: three or so practical sessions (hopefully we have time)
  - There are two seminars! **You come to just one of the two sessions**, either Wednesdays (Group 01) or Thursdays (Group 02)
Registration:

- If you are a CIS Student: check whether you are registered for *both* the Vorlesung and the Seminar (these are **two things** in LSF!)
  - Be sure to ignore the Modulteilprüfung entries, these will be deleted
- There are a good number of people only in the Vorlesung
- There are a few people only in the Seminar
- With respect to the Seminar, for everyone:
  - Most of you are registered in LSF for both groups, **PLEASE DROP ONE GROUP**!
  - If both slots are OK, please select Thursday, your colleagues will thank you
Information Extraction – Administravia - III

- Vorlesung and Seminar are two separate courses (in same module for CIS people)
  - However, there may be some shifting around of slots depending on time constraints
- Vorlesung (Grade):
  - Klausur (possibly 08.02, no discussion of this today please)
- Seminar (Grade):
  - Referat
  - Hausarbeit (write-up of the Referat) (6 pages, due 3 weeks after you hold your Referat)
  - The Hausarbeit can also include practical exercises (optional, extra points)
- CIS-ler: No Notenverbesserung (everyone else: ask in your Fachschaft!)
Information Extraction - Administravia - IV

• First Seminar
  • I will collect information on who you are and your interests
  • And I want to know what you want to learn in this class!
Information Extraction – Administravia - V

- Syllabus: updated dynamically on my web page (see also WS last year, but there will be some differences)
  - Brief idea at end of this slide deck (if we finish, then today)
- List of Referatsthemen
  - This will be presented soon in the Seminar, probably in two weeks
- Literature:
    - Please read the introduction for next week (it is available on the web page!)

Please note: The slides contain a mistake in the first bullet point. It should read: "Syllabus: updated dynamically on my web page (see also WS last year, but there will be some differences)."
Information Extraction - Administravia - VI

• There will also be guest lectures from Dr. Fabienne Braune, Matthias Huck, Dr. Tsuyoshi Okita
• Our tutor, Fabian Dreer, will help with the exercises and available to help you with any questions
• Questions?
Information Extraction

• An introduction to the course
  • The topic "Information Extraction" means different things to different people
  • In this course we will look at several different perspectives
  • There is unfortunately no comprehensive textbook that includes all of these perspectives
My Biases

• As you may have noticed by now: I am from the US (PhD in Computer Science from USC/ISI AI division)
• I am on permanent staff here at CIS
• I do research in the broad area of statistical NLP
  • I mostly work on machine translation, and related structured prediction problems (e.g., treebank-based syntactic parsing, generation using sequence (tagging) models)
  • I also work on other multilingual problems such as cross-language information retrieval
• With respect to rule-based NLP (with manually written rules), I'll try to be as fair as humanly possible, I do use these techniques sometimes too
Outline for today

• Motivation
  • Problems requiring information extraction
  • Basic idea of the output
• Abstract idea of the core of an information extraction pipeline
• Course topics
A problem

Mt. Baker, the school district

Baker Hostetler, the company

Baker, a job opening

Slide from Cohen/Mccallum
Bakery Jobs on CareerBuilder.com
www.careerbuilder.com/jobs/keyword/bakery

Baker Jobs, Employment | Indeed.com
www.indeed.com/q-Baker-jobs.html
Jobs 1 - 10 of 16047 – 16047 Baker Jobs available on Indeed.com. one search, all jobs.

Job Openings - Baker University
www.bakeru.edu/jobs
If you are seeking employment in any of these areas, contact Baker University.

Baker, LA Jobs on CareerBuilder.com
www.careerbuilder.com/Jobs/Baker/
Jobs 1 - 25 of 948 – Looking for Baker, LA Jobs? See currently available job openings on CareerBuilder.com. Browse the current listings and fill out job ...

Down Under Bakery Pies: Job Openings at DUB Pies
www.dubpies.com/jobs.php
Listing of job openings at DUB Pies. Down Under Bakery (DUB) Pies is looking for more staff - check out our list of vacancies.

Field Engineers | Geoscience | Jobs and Careers at Baker Hughes
jobs.bakerhughes.com/
... Oil and Natural Gas? Baker Hughes has career information for you on these, more. ... Search Jobs. Baker Hughes Jobs ... Recent Job Openings. Completion ...

Corner Bakery Job Openings | Glassdoor
www.glassdoor.com/Job/Corner-Bakery-Job-Openings-E297310_P2...
46 Corner Bakery job openings. Search job openings, see if they fit - company salaries, reviews, and more posted by Corner Bakery employees.

Jobs - Baker University
www.bakeru.edu/jobs
See links at left for a complete list of Baker University job openings. It is the policy of Baker University to afford equal opportunity for all persons without distinction ...
A solution
### Job Openings:

**Category:** Food Services  
**Keyword:** Baker  
**Location:** Continental U.S.

<table>
<thead>
<tr>
<th>Job Title</th>
<th>Company</th>
<th>Location</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food Pantry Workers</td>
<td>Lutheran Social Services</td>
<td>Archbold, OH</td>
<td>October 11, 2002</td>
</tr>
<tr>
<td>Cooks</td>
<td>Lutheran Social Services</td>
<td>Archbold, OH</td>
<td>October 11, 2002</td>
</tr>
<tr>
<td>Bakers Assistants</td>
<td>Fine Catering by Russell Morin</td>
<td>Attleboro, MA</td>
<td>October 11, 2002</td>
</tr>
<tr>
<td>Baker’s Helper</td>
<td>Bird-in-Hand</td>
<td>United States</td>
<td>October 11, 2002</td>
</tr>
<tr>
<td>Assistant Baker</td>
<td>Gourmet To Go</td>
<td>Maryland Heights, MO</td>
<td>October 10, 2002</td>
</tr>
<tr>
<td>Cake Decorator/Baker</td>
<td>Mandalay Bay Hotel and Casino</td>
<td>Las Vegas, NV</td>
<td>October 08, 2002</td>
</tr>
<tr>
<td>Shift Supervisors</td>
<td>Brueggers Bagels</td>
<td>Minneapolis, MN</td>
<td>October 08, 2002</td>
</tr>
</tbody>
</table>
Title: Ice Cream Guru

Description: If you dream of cold creamy…

Contact: susan@foodscience.com

Category: Travel/Hospitality

Function: Food Services
Another Problem
Often structured information in text

0.44 CT ROUND CUT DIAMOND PENDANT 14 K WHITE GOLD
Classic style and beauty, this comfortable 14 K White gold pendant contains:
An Ideal cut Round 0.44 CT Diamond, in a magnificent high polish bezel.
- Color: F
- Clarity: SI-1
- Setting: 14 K White Gold
- Chain: 16 inches 14 K White Gold
- Weight: 3.4 g
- Measurements: 10 mm x 10 mm
- Retail Price: $2,319.00
- Close Out Price: $889.00


Thirteenth International Conference on Computational Linguistics, Volume 3

Slide from Cohen/McCallum
Information Extraction (IE) is the process of extracting structured information (e.g., database tables) from unstructured machine-readable documents (e.g., Web documents).

Information Extraction

<table>
<thead>
<tr>
<th>GName</th>
<th>FName</th>
<th>Occupation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elvis</td>
<td>Presley</td>
<td>singer</td>
</tr>
<tr>
<td>Elvis</td>
<td>Hunter</td>
<td>painter</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

“Seeing the Web as a table”

Elvis Presley was a famous rock singer.

Mary once remarked that the only attractive thing about the painter Elvis Hunter was his first name.
Defining an IE problem

• In what I will refer to as "classic" IE, we are converting documents to one or more table entries
  • There are other kinds of IE, we will talk about those later
• The design of these tables is usually determined by some business need
• Let's look at the table entries for a similar set of examples to the ones we just saw
## Motivating Examples

### Table: Job Titles, Types, and Locations

<table>
<thead>
<tr>
<th>Title</th>
<th>Type</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business strategy Associate</td>
<td>Part time</td>
<td>Palo Alto, CA</td>
</tr>
<tr>
<td>Registered Nurse</td>
<td>Full time</td>
<td>Los Angeles</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

### Image: Job Search Results

- **RN-Registered Nurse/LVN-Licensed Vocational Nurse**
  - Job type: Full-Time/Part-Time
  - Maxim's office in Sherman Oaks is seeking compassionate Registered Nurses (RN) and Licensed Vocational Nurses (LVN)...

- **Nurse Practitioner - Acute Care Nurse Practitioner**
  - Job type: Full-Time
  - Vanderbilt University Medical Center is currently hiring Nurse Practitioners to join our team...

---

*Slide from Suchanek*
### Motivating Examples

<table>
<thead>
<tr>
<th>Name</th>
<th>Birthplace</th>
<th>Birthdate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elvis Presley</td>
<td>Tupelo, MI</td>
<td>1935-01-08</td>
</tr>
</tbody>
</table>

**Biography for Elvis Presley**

**Name:** Elvis Presley

**Birthplace:** Tupelo, MI

**Birthdate:** 1935-01-08

**Date of Birth:** 8 January 1935, Tupelo, Mississippi, USA

**Date of Death:** 16 August 1977, Memphis, Tennessee, USA (cardiac arrhythmia)

**Height:** 6' (1.83 m)

**Mini Biography**

Elvis Aaron Presley, in the humblest of circumstances, was born to Vernon and Gladys Presley in a two-room house in Tupelo, Mississippi on January 8, 1935. His twin brother, Jessie Garon, was stillborn, leaving Elvis to grow up as an only child. He and his parents moved to Memphis, Tennessee in 1948, and Elvis graduated from Humes High School there in 1953.
Motivating Examples

Information Extraction: Techniques and Challenges

Ralph Grishman

Information Integration Papers


<table>
<thead>
<tr>
<th>Author</th>
<th>Publication</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grishman</td>
<td>Information Extraction...</td>
<td>2006</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Slide from Suchanek
### Motivating Examples

<table>
<thead>
<tr>
<th>Product</th>
<th>Type</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dynex 32&quot;</td>
<td>LCD TV</td>
<td>$1000</td>
</tr>
<tr>
<td>Ballroom Dance Shoe</td>
<td>1 new from $49.95</td>
<td></td>
</tr>
<tr>
<td>X-Strap Bag</td>
<td>1 new from</td>
<td></td>
</tr>
<tr>
<td>Dynex™ - 32&quot; Class / 720p / 60Hz / LCD HDTV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dynex™ - 24&quot; Class / 1080p / 60Hz / LCD HDTV</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Slide from Suchanek*
Information Extraction (IE) is the process of extracting structured information from unstructured machine-readable documents.
Information Extraction

**Traditional definition:** Recovering structured data from text

**What are some of the sub-problems/challenges?**

--

Slide from Nigam/Cohen/McCallum
Information Extraction?

- Recovering structured data from text
- Identifying fields (e.g. named entity recognition)

Board Members

- Itzhak Fisher
  - Chairman of Nielsen BuzzMetrics

- Thom Mastrelli
  - Executive Vice President, Corporate Development, VNU

- Jonathan Carson
  - CEO of Nielsen BuzzMetrics

- Mahendra Vora
  - CEO and Owner, Vora Technology Park

- Ori Levy
  - President of Nielsen BuzzMetrics
  - Israel

- Ron Schneier
  - Senior Vice President and General Manager, Nielsen Ventures

- James O’Hara
  - Senior Vice President and Chief Financial Officer, VNU’s Media Measurement and Information Group

Slide from Nigam/Cohen/McCallum
Information Extraction?

- Recovering structured data from text
  - Identifying fields (e.g. named entity recognition)
  - Understanding relations between fields (e.g. record association)
Information Extraction?

• Recovering structured data from text
  • Identifying fields (e.g. named entity recognition)
  • Understanding relations between fields (e.g. record association)
  • Normalization and deduplication
Information extraction

• Input: Text Document
  • Various sources: web, e-mail, journals, ...
• Output: Relevant fragments of text and relations possibly to be processed later in some automated way

Slide from McCallum
Not all documents are created equal...

- Varying regularity in document collections
- Natural or unstructured
  - Little obvious structural information
- Partially structured
  - Contain some canonical formatting
- Highly structured
  - Often, automatically generated
BACKGROUND: The most challenging aspect of revision hip surgery is the management of bone loss. A reliable and valid measure of bone loss is important since it will aid in future studies of hip revisions and in preoperative planning. We developed a measure of femoral and acetabular bone loss associated with failed total hip arthroplasty. The purpose of the present study was to measure the reliability and the intraoperative validity of this measure and to determine how it may be useful in preoperative planning. METHODS: From July 1997 to December 1998, forty-five consecutive patients with a failed hip prosthesis in need of revision surgery were prospectively followed. Three general orthopaedic surgeons were taught the radiographic classification system, and two of them classified standardized preoperative anteroposterior and lateral hip radiographs with use of the system. Interobserver testing was carried out in a blinded fashion. These results were then compared with the intraoperative findings of the third surgeon, who was blinded to the preoperative ratings. Kappa statistics (unweighted and weighted) were used to assess correlation. Interobserver reliability was assessed by examining the agreement between the two preoperative raters. Prognostic validity was assessed by examining the agreement between the assessment by either Rater 1 or Rater 2 and the intraoperative assessment (reference standard). RESULTS: With regard to the assessments of both the femur and the acetabulum, there was significant agreement (p < 0.0001) between the preoperative raters (reliability), with weighted kappa values of >0.75. There was also significant agreement (p < 0.0001) between each rater’s assessment and the intraoperative findings (validity). CONCLUSIONS: With use of the newly developed classification system, preoperative radiographs are reliable and valid for assessment of the severity of bone loss that will be found intraoperatively.
Partially Structured: Seminar Announcements

Extract time, location, speaker, etc.

We will finish the CSE AI research seminar this Monday, November 26th, with speaker Dave Kauchak from the UCSD AI lab. We meet in AP&M 4982 at 12:10PM. Free pizza!

Title:

Boosting for information extraction

Abstract:

In this talk I will examine Boosted Wrapper Induction (BWI, Freitag & Kushmerick) as an exemplar of recent rule-based information extraction (IE) techniques. Results will be shown for BWI on a wide variety of tasks than has previously been studied, including several natural text document collections. I will examine these results and show how the tests performed allow for a systematic analysis of how each of BWI's algorithmic components, particularly boosting, contributes to its performance over comparable methods. I will also present a new metric, the SWL-Ratio, which is a quantitative measure of the regularity of an extraction task, and its use in evaluating extraction tasks.

First up is Gary Cottrell.

A Neural Network that Perceives and Categorizes Facial Expressions

Abstract:

How do we perceive emotions in facial expressions? On the one hand, findings show that we map facial expressions into discrete categories, as in color and phoneme perception, with sharp boundaries between emotions and better discrimination between pairs of stimuli that straddle a category boundary. On the other hand, there is good evidence
Highly Structured: Zagat’s Reviews

Extract restaurant, location, cost, etc.
Landscape of IE Tasks: Document Formatting

Text paragraphs without formatting

Astro Teller is the CEO and co-founder of BodyMedia. Astro holds a Ph.D. in Artificial Intelligence from Carnegie Mellon University, where he was inducted as a national Hertz fellow. His M.S. in symbolic and heuristic computation and B.S. in computer science are from Stanford University.

Non-grammatical snippets, rich formatting & links

Barto, Andrew G. (413) 545-2109 barto@cs.umass.edu CS276
Professor.
Computational neuroscience, reinforcement learning, adaptive motor control, artificial neural networks, adaptive and learning control, motor development.

Berger, Emery D. (413) 577-4211 emery@cs.umass.edu CS344
Assistant Professor.

Brooke, Oliver (413) 577-0334 oli@cs.umass.edu CS246
Assistant Professor.

Clarke, Lori A. (413) 545-1328 clarke@cs.umass.edu CS304
Professor.
Software verification, testing, and analysis; software architecture and design.

Grammatical sentences and some formatting & links

Dr. Steven Minton - Founder/CTO
Dr. Minton is a fellow of the American Association of Artificial Intelligence and was the founder of the Journal of Artificial Intelligence Research. Prior to founding Fetch, Minton was a faculty member at USC and a project leader at USC's Information Sciences Institute. A graduate of Yale University and Carnegie Mellon University, Minton has been a Principal Investigator at NASA Ames and taught at Stanford, UC Berkeley and USC.

Tables

8:30 - 9:30 AM Invited Talk: Plausibility Measures: A General Approach for Representing Uncertainty
Joseph Y. Halpern, Cornell University
9:30 - 10:00 AM Coffee Break
10:00 - 11:30 AM Technical Paper Sessions:

Cognitive Robotics
739: A Logical Account of Causal and Topological Maps with
Emma Romanillo and Benjamin Kuipers

Logic Programming
116: A System for Solving Abduction Problems with
Marc Demuynck, Antoninus Kakas, and Bert Van Nuffelen

Natural Language Generation
758: Title Generation for Machine-Translated Documents with
Rong Jin and Alexander G. Hauptmann

Complexity Analysis
417: Let's go Nuts: Complexity of Nested Circumscription and Abnormality Theories with
Marco Ciardelli, Thomas Eiter, and Georg Gottlob

Neural Networks
179: Knowledge Extraction and Comparison from Local Function Networks with
Kenna McCarthy, Stefan Werner, and John MacIntyre

Games
71: Iterative Widening
D. Cazenave

549: Online-Execution of Co-Opting Plans with
Hendrik Grosskreutz and Gerhard Lakemeyer

131: A Comparative Study of Logic Programming with
Preference Tuerkenn Schaub and Kewen

246: Dealing with Dependencies between Content Planning and Surface Realization in a Pipeline Generation with

470: A Perspective on Knowledge Compilation with
Dana and Pierre Marquis

258: Violations-Guided Learning for Constrained Formulations in Neural-Network Time-series Prediction with

353: Temporal Difference Learning
J. Bagnell

Slide from McCallum
Landscape of IE Tasks
Intended Breadth of Coverage

Web site specific
Formatting
Amazon.com Book Pages

Genre specific
Layout
Resumes

Wide, non-specific
Language
University Names

Slide from McCallum
Landscape of IE Tasks: Complexity of entities/relations

**Closed set**
- U.S. states
  - He was born in **Alabama**…
  - The big **Wyoming** sky…

**Regular set**
- U.S. phone numbers
  - Phone: (413) 545-1323
  - The CALD main office is **412-268-1299**

**Complex pattern**
- U.S. postal addresses
  - University of Arkansas
    - P.O. Box **140**
    - Hope, AR
  - Headquarters:
    - 1128 Main Street, 4th Floor
    - Cincinnati, Ohio 45210

**Ambiguous patterns, needing context and many sources of evidence**
- Person names
  - …was among the six houses sold by **Hope Feldman** that year.
  - Pawel Opalinski, Software Engineer at WhizBang Labs.
Jack Welch will retire as CEO of General Electric tomorrow. The top role at the Connecticut company will be filled by Jeffrey Immelt.

<table>
<thead>
<tr>
<th>Single entity</th>
<th>Binary relationship</th>
<th>N-ary record</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Person</strong>: Jack Welch</td>
<td><strong>Relation</strong>: Person-Title</td>
<td><strong>Relation</strong>: Succession</td>
</tr>
<tr>
<td><strong>Person</strong>: Jeffrey Immelt</td>
<td><strong>Person</strong>: Jack Welch</td>
<td><strong>Company</strong>: General Electric</td>
</tr>
<tr>
<td><strong>Title</strong>: CEO</td>
<td><strong>Title</strong>: CEO</td>
<td><strong>Out</strong>: Jack Welch</td>
</tr>
<tr>
<td><strong>Location</strong>: Connecticut</td>
<td><strong>Company</strong>: General Electric</td>
<td><strong>In</strong>: Jeffrey Immelt</td>
</tr>
<tr>
<td><strong>Location</strong>: Connecticut</td>
<td><strong>Location</strong>: Connecticut</td>
<td><strong>Location</strong>: Connecticut</td>
</tr>
</tbody>
</table>

"Named entity" extraction
Association task = Relation Extraction

• Checking if groupings of entities are instances of a relation

1. Manually engineered rules
   • Rules defined over words/entities: “<company> located in <location>”
   • Rules defined over parsed text:
     • “((Obj <company>) (Verb located) (*) (Subj <location>))”

2. Machine Learning-based
   • Supervised: Learn relation classifier from examples
   • Partially-supervised: bootstrap rules/patterns from “seed” examples
May 19 1995, Atlanta -- The Centers for Disease Control and Prevention, which is in the front line of the world's response to the deadly Ebola epidemic in Zaire, is finding itself hard pressed to cope with the crisis...

<table>
<thead>
<tr>
<th>Date</th>
<th>Disease Name</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan. 1995</td>
<td>Malaria</td>
<td>Ethiopia</td>
</tr>
<tr>
<td>July 1995</td>
<td>Mad Cow Disease</td>
<td>U.K.</td>
</tr>
<tr>
<td>Feb. 1995</td>
<td>Pneumonia</td>
<td>U.S.</td>
</tr>
</tbody>
</table>
We show that CBF-A and CBF-C interact with each other to form a CBF-A-CBF-C complex and that CBF-B does not interact with CBF-A or CBF-C individually but that it associates with the CBF-A-CBF-C complex.

CBF-A interact complex CBF-C

CBF-B associates CBF-A-CBF-C complex
John Fitzgerald Kennedy was born at 83 Beals Street in Brookline, Massachusetts on Tuesday, May 29, 1917, at 3:00 pm,[7] the second son of Joseph P. Kennedy, Sr., and Rose Fitzgerald; Rose, in turn, was the eldest child of John "Honey Fitz" Fitzgerald, a prominent Boston political figure who was the city's mayor and a three-term member of Congress. Kennedy lived in Brookline for ten years and attended Edward Devotion School, Noble and Greenough Lower School, and the Dexter School, through 4th grade. In 1927, the family moved to 5040 Independence Avenue in Bronx, New York City; two years later, they moved to 294 Pondfield Road in Bronxville, New York, where Kennedy was a member of Scout Troop 2 (and was the first Boy Scout to become President).[8] Kennedy spent summers with his family at their home in Hyannisport, Massachusetts, and Christmas and Easter holidays with his family at their winter home in Palm Beach, Florida. For the 5th through 7th grade, Kennedy attended Riverdale Country School, a private school for boys. For 8th grade in September 1930, the 13-year old Kennedy attended Canterbury School in New Milford, Connecticut.
Rough Accuracy of Information Extraction

<table>
<thead>
<tr>
<th>Information type</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entities</td>
<td>90-98%</td>
</tr>
<tr>
<td>Attributes</td>
<td>80%</td>
</tr>
<tr>
<td>Relations</td>
<td>60-70%</td>
</tr>
<tr>
<td>Events</td>
<td>50-60%</td>
</tr>
</tbody>
</table>

- Errors cascade (error in entity tag $\rightarrow$ error in relation extraction)
- These are very rough, actually optimistic, numbers
  - Hold for well-established tasks, but lower for many specific/novel IE tasks
What we will cover in this class (briefly)

- History of IE, Related Fields
- Source Selection
- Tokenization and Normalization
- Named Entity Recognition (NER)
- Several topics in general machine learning (applied to NER)
- Instance Extraction
- Fact/Event Extraction
- Ontological IE/Open IE
Seminar

• Please attend the seminar tomorrow if you did not attend it today, will be about 20 minutes, I am collecting information that is very useful for planning

• Also, don't forget the reading for next week!

• Sarawagi: Information Extraction (available from web page). Read the introduction!
• Thank you for your attention!