"Unstructured information accounts for more than 70%-80% of all data in organizations and is growing 10-50x more than structured data"
What can we do with unstructured textual data?

- Help answer questions
- Detect if customer opinions are positive or negative
- Efficiently search large amount of documents (litigation and patent search)
- Help provide better services for public health (e.g. find patients for Clinical trials, help in epidemiologic studies)
- Detect risks
- Help match candidates and job offers.

"Unstructured information accounts for more than 70%-80% of all data in organizations and is growing 10-50x more than structured data."

[Diagram showing various categories like Logistics, Delivery, Directions, Specifications, etc.]
How do we make computers understand unstructured text?

Natural Language Processing (NLP) helps address this question.
Xerox NLP technology

• More than 30 years of R&D in Natural Language Processing technologies awarded in many international competitions

• We master every single step in the processing chain (from text to semantics) → it allows us to fine tune our technology for specific applications and to adapt to new domains

• Multiple languages (English, French, German, Italian, Spanish, Portuguese, Dutch)
How it works

“Journalists went to Xerox on October, 3rd.”

DEMO Parsing

Xerox Natural Language Processing (NLP) technologies in applications

Xerox NLP tools

Opinion Mining
Sentiment analysis

Customer Relationship Management

Customer reviews
Blogs
Forums

Redaction
Evidence searching

Litigation

Emails

Anonymization
Risk detection

HEALTH-CARE

Patient data
Healthcare in numbers

- Healthcare is the 1st job provider in Europe
- The number of people over 65 will double in 2050
- In 2050 health spending will be 10-13% of Gross Domestic Product

Key Areas
- Predicting disease
- Empowering patients
- Reduction of risks
- Health Information Exchange

European Commission 7th Framework Program Injected 2 Billion € to boost Information & Communication Technologies
eHealth: a global trend toward Health Information Exchange

- Move from paper to digital
- Organize & structure information (e.g. database)
- Develop & use information encoding standards
- Develop system interoperability for global health information exchange
- Emergence of new services (monitoring / diagnosis / prognosis)

**Investments** to drive down these costs while improving quality of service and patient safety

Percentage of General Practitioners using a computer (Random sample of 6789 GPs in 29 countries)

Sources: empirica ICT and eHealth use among GPs in Europe 2007. Bonn, April 2008
Xerox Europe research in healthcare

• Mine scientific literature to build a gene-protein interaction network
• Bridge the gap between research and clinical care
• Capture patient health statistics from hospital reports
• Monitor Hospital Acquired Infections from hospital reports
A Hospital Acquired Infection is:
“An infection occurring in a hospital that was not present or incubating at the time of admission. A period of 48 hours (or corresponding to the incubation period) is considered to characterize HAI. For surgery a period of 30 days is considered and extended to 12 months in case of implanted device.”

- 5% to 10% of Hospitalized patients will get an HAI
- In Europe, 3 Million cases / 50 000 deaths a year

What is being done to reduce the risk
- Watch groups
- Training & Guidelines
- Process & Monitoring

A French funded research project to monitor risks inside patient discharge summaries
How to address the problem

- Search for facts/events (symptoms, treatments, bacteria, …)
- Context and sense making (e.g. having fever vs. no fever)
- Chronology of events (incubation period > 2 days)
- Find relations between events
- Flexible decision rules (unwritten knowledge)
- Alert

“…The postoperative consequences were marked by abdominal pain and fever, associated with a hyperleucocytosis and inflammation. It was due to multiple intra-peritoneal abscesses that required a peritoneal toilet on September 29th. It was an infection with Klebsiella only sensitive to Tienam …”
1st task: Enforcing privacy

<table>
<thead>
<tr>
<th>Texte source</th>
<th>Texte anonymisé</th>
</tr>
</thead>
<tbody>
<tr>
<td>Service d'Anesthésie-Réanimation</td>
<td>Service d'Anesthésie-Réanimation</td>
</tr>
<tr>
<td>Professeur J.CLOCHE</td>
<td>Professeur [PERSONNE = MEDECIN]</td>
</tr>
<tr>
<td>Praticiens Hospitaliers:</td>
<td>Praticiens Hospitaliers:</td>
</tr>
<tr>
<td>CERISSE P.A (04 78 12 35 11)</td>
<td>[PERSONNE = AUTRE] (TELFAI)</td>
</tr>
<tr>
<td>E.mail : <a href="mailto:pierre-antione.Cerisse@chu-lyon.fr">pierre-antione.Cerisse@chu-lyon.fr</a></td>
<td>E.mail : [EMAIL]</td>
</tr>
<tr>
<td>COMPTE RENDU D'HOSPITALISATION</td>
<td>COMPTE RENDU D'HOSPITALISATION</td>
</tr>
<tr>
<td>Né(e) le 08/12/1935</td>
<td>Né(e) le [AGE_PATIENT = 72]</td>
</tr>
<tr>
<td>Adressé au Professeur MER chls</td>
<td>Adressé au Professeur [PERSONNE = MEDECIN] [LIEU]</td>
</tr>
<tr>
<td>Histoire de la maladie:</td>
<td>Histoire de la maladie:</td>
</tr>
<tr>
<td>Monsieur Durand a été opéré par le</td>
<td>Monsieur [PERSONNE = AUTRE] a été opéré par le</td>
</tr>
<tr>
<td>Professeur MER d'une gastrectomie totale pour adénocarcinome.</td>
<td>Professeur [PERSONNE = MEDECIN] d'une gastrectomie totale pour adénocarcinome.</td>
</tr>
<tr>
<td>Dans les suites post-opératoires, il est découvert une fistule de l'anastomose œsophago-jéjunale qui conduit à la mise en place d'une endoprothèse œsophagienne le 04/09/2008. Devant la persistance des symptômes, un scanner est réalisé le 08/09/2008 et retrouve une volumineuse collection hydroaérique abdominale.</td>
<td>Dans les suites post-opératoires, il est découvert une fistule de l'anastomose œsophago-jéjunale qui conduit à la mise en place d'une endoprothèse œsophagienne le [T-5J]. Devant la persistance des symptômes, un scanner est réalisé le [T-1J] et retrouve une volumineuse collection hydroaérique abdominale.</td>
</tr>
</tbody>
</table>

Same linguistic technology removes names and all personal data from texts
Evaluation at the end of the project

Objective for HAI experts was to miss as few cases as possible
- 815 patient discharge summaries from
  - Intensive Care Unit
  - Stomach Surgery
  - Knee Surgery
  - Brain Surgery

- 93% of problems detected
- System prediction was accurate 87% of the time
Going Live in Lyon Hospital (CHU)

Hospices Civils de Lyon
Annual activity

• 908,619 consultations
• 1,469,290 days of hospitalisation
• 219,354 emergency admissions
• 9.9 million laboratory procedures