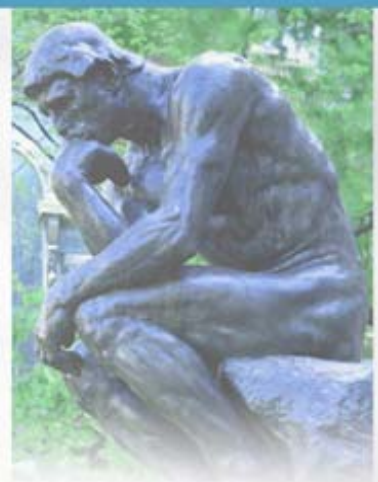
"Integration of NLP technology into KM can save people's lives"



Langtech 2003 - Paris

Maarten Laga

CEO

Language and Computing

Language Computing

Language and Computing

Agenda

- Introduction
- A problem statement
- How NLP helps
- NLP technological requirements
- Conclusion

Language and Computing

Introduction

- NLP as research domain
- NLP as developping technology
- NLP as nice-to-have technology

- NLP as NEED-to-have technology
  - Proven ROI in relevant domains

Language and Computing

Problems facing healthcare environments :

- Medication errors
- Inefficient treatments
- No cure for certain diseases
- Bioterrorism
- Long drug development time

To a large extend, these problems can all be caused by incomplete, non-available or contradictory information !!

L&C
Language and Computing

Healthcare organizations have lots of information, but most of it ...

- Is stored at different fysical locations
- Is stored in an unstructured format (free text)
- Is not immediately available when needed
- Cannot be found with traditional technologies
- Is stored using different terminologies to mean the same thing
- Is not analysed to discover trends or threats

Language and Computing

THE RESULT ?

The number of deaths due to medical errors in hospitals is estimated at 44,000 to 98,000 annually in the USA

(Institute of Medicine, USA, 1999)

Language and Computing

How do healthcare organizations aim to resolve these problems ?

- Electronic Medical Records (EMR) aim to collect all patient-related information and make it available for everyone from anywhere (e.g. The NHS in the UK)
- Clinical databases are linked: e.g. drugs, medications
- Technologies such as search and retrieval and data mining are applied to large clinical document sets to search for or extract valuable information.

L&C
Language and Computing

The EMR as the center of the Healthcare process

Gartner

L&C
Language and Computing

BUT …

All these systems still contain lots of free-text information, that cannot be processed by computers

What information is present in unstructured clinical documents ?

- Some crucial details
- The information feeding clinical decisions
- What the author really thought
- The real story

Language and Computing

Healthcare IT systems can't read it !

For <u>most</u> computer applications

Text is

- Messy
- Ambiguous
- Disorderly
- Incomplete
- Ill behaved
- Ignored

But,...for a (human) clinician

Text is

- Expressive
- Informative
- Valuable
- Costly to ignore
- Expensive to replace
- Here to stay

L&C
Language and Computing

Medical NLP

To provide patients with the best possible healthcare, Healthcare IT systems need to bring the best possible information to the clinical decision makers at the right time.

To be able to do this, *computers need a full and detailed understanding* of medical natural language.

Medical Natural Language Processing technology delivers this understanding.

Language and Computing

NLP requirements in this context

Extract meaning from free text :

- Concepts mentioned
- Context of concepts
  - » Negation
  - » Modality
  - » Document Section
  - » Parsing the relations to other concepts discussed
- Relation of explicit concepts to implicit knowledge
- Ranking the most important concepts
- Understanding the propositions expressed about the concepts.

Language and Computing

The L&C approach:

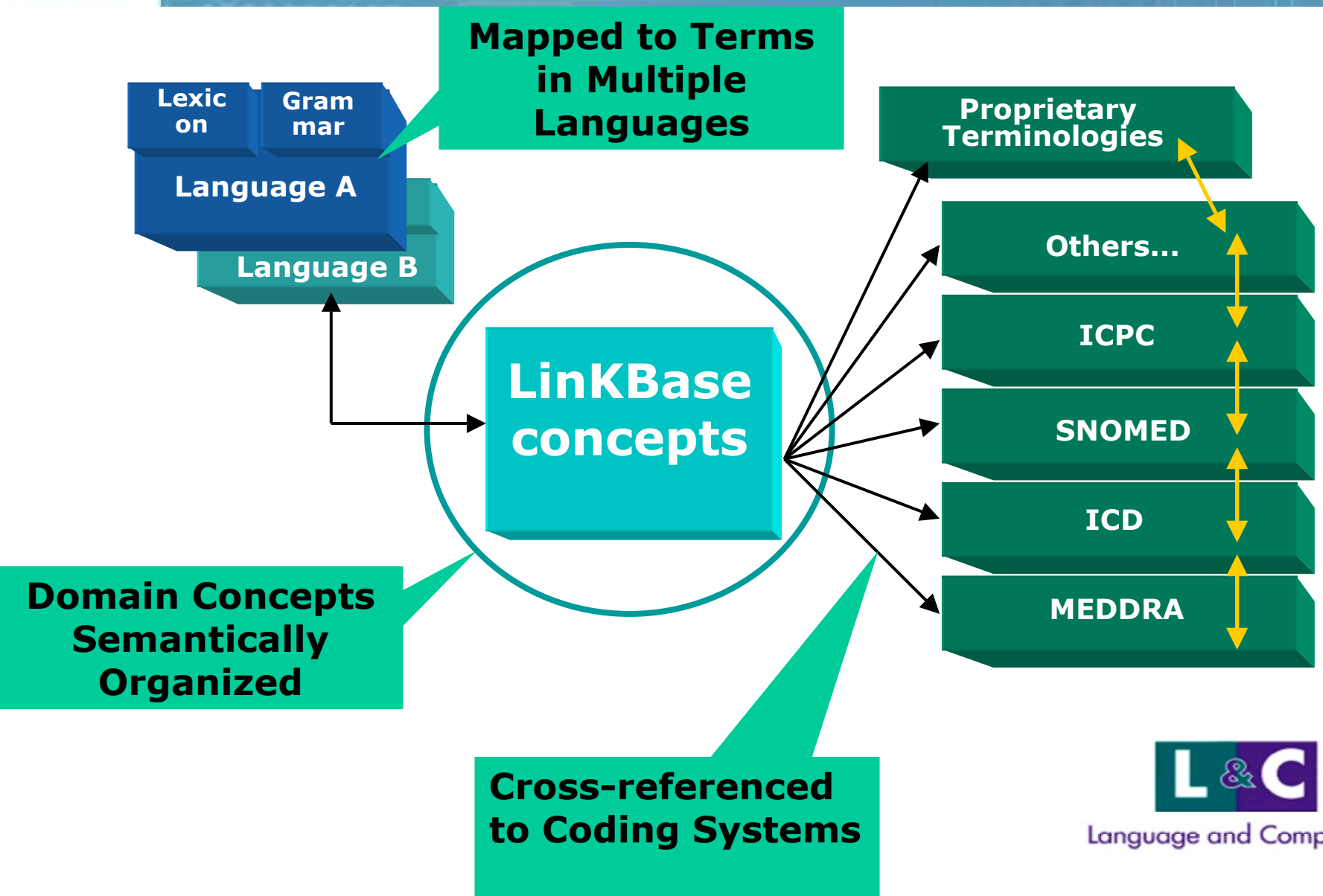
Two major components:

- ✓ NLP technology:
  - software for language analysis, semantic indexing, information retrieval, fact extraction

- ✓ Formal ontology:
  - machine understandable knowledge base of a given domain (context information)

Language and Computing

LinKBase® Medical Ontology



Language and Computing

LinKBase® Medical Ontology

- **Contains more than 1.5 million medical concepts**
- **540 different linktypes generate 3 million relations between the concepts**
- **3 million terms in 5 languages (English, French, Spanish, Dutch, Italian + more)**

LinKBase delivers 7.5 million knowledge elements making it the largest medical knowledge base in the world

L&C
Language and Computing

Examples

- Processing of information residing in Electronic Medical Records

- Medical Information Extraction

Language and Computing

EMR Example

PRESENTING COMPLAINT

The patient is a 36-year-old woman who presents for an evaluation of her life-long asthma. Her asthma has been poorly controlled since childhood; her symptoms seemed to worsen over the past three years. Over that period she has wheezed daily, had numerous emergency department visits, required monthly visits to her physician and required daily prednisone. Her symptoms were minimal at night and were rarely exacerbated by exercise. The patient was particularly concerned about longer recovery periods between attacks. These attacks required hospitalization four times in the five months preceding referral. She was diagnosed with radiologically confirmed pneumonia three times in the year prior to referral. The patient was evaluated at <medical center> to confirm her diagnosis of asthma and to evaluate for atopy and other allergic triggers.

Past medical history is significant for eczema. The patient reported allergies since childhood, with anaphylactic reactions to specific foods including nuts, fish and raw eggs. There is no history of aspirin sensitivity or nasal polyposis. She also has a history of gastroesophageal reflux disease diagnosed by endoscopy. She suffers from osteopenia and cataracts related to her chronic corticosteroid use.

Language and Computing

EMR Example

Understands indirect references to patient

PRESENTING COMPLAINT .

The patient is a 36-year-old ***woman*** ***who*** (Antecedent) woman presents for an evaluation of ***her*** (Antecedent) patient life-long asthma . ***Her*** (Antecedent) patient asthma has been poorly ***controlled*** since childhood ; ***her*** (Antecedent) patient ***symptoms*** seemed to worsen over the past three years . Over that ***period*** ***she*** (Antecedent) patient has wheezed daily , had numerous emergency department visits , daily requirement monthly visits to ***her*** (Antecedent) patient physician and daily requirement daily prednisone (Medication) . ***Her*** (Antecedent) patient ***symptoms*** were minimal at night and were rarely exacerbated by exercise . The patient (rewritten: patient concerned patient) was particularly concerned about longer recovery ***periods*** between ***attacks*** . These ***attacks*** required hospitaliza[illegible]n four times in the five ***months preceding*** referral . ***She*** (Antecedent) patient was ***diagnosed*** with radiologically confirmed [illegible]monia three times in the year prior to referral . The patient (rewritten: patient evaluation patient) was evaluated at < medica[illegible] to confirm ***her*** (Antecedent) patient diagnosis of asthma and to evaluate for atopy and ***other*** allergic triggers .

Combines text words to identify complex concepts

Past medical history is significant for ec[illegible]edication) since childhood , with anaphylactic reactions to specific foods ***including*** nuts , ***fish*** and raw eggs[illegible]egg raw raw eggs) . There is no ***history*** of aspirin (Medication) ***sensitivity*** or nasal polyposis (Neg) . She also has a ***history*** of gastroesophageal reflux disease ***diagnosed*** by endoscopy . ***She*** (Antecedent) She suffers from osteopenia and cataracts ***related*** to ***her*** (Antecedent) She chronic corticosteroid use .

EMR Example

Bronchoscopy revealed friable , irritated mucosa and normal (rewritten: Bronchoscopy normal normal normal mucosa normal) left sided (rewritten: side) airways . The right middle lobe orifice was quite ***small*** , and
significantly deformed and was only able to be entered with a ***deep inspira***
right middle lobe revealed normal cell ***count*** and differential ; however elec
Langerhan ? s type macrophages with pleomorphic structures noted to be a ached to them (Antecedent) macrophages . These are ultrastructurally consistent with Mycoplasma pneumoniae .

domain: BRONCHOSCOPY NORMAL
MEDDRA: MEDDRA : 10006481 : BRONCHOSCOPY NORMAL
SNOMED-CT: SNOMED-CT : 165142008 : BRONCHOSCOPY NORMAL (FINDING)

IMPRESSION AND ***PLAN*** Bronchomalacia , acquired . Possibly a ***res*** congenital weakness of the fibrous tissue ***combined*** with chronic inflammatory irritation (Mod) . Distal and ed . Recommend ***combination*** of flutter valve use , pursed ***lip*** breathing during exercise and nasal (rewritten: nasal us CPAP to wean from corticosteroids (Medication) and asthma medications (rewritten: asthma medications asthma medi cations) . Surgical treatment strategy might be indicated (Mod) . Consider surgical (Medication) stent placement to way collapse .

Combines text words across the sentence to identify a complex concept.

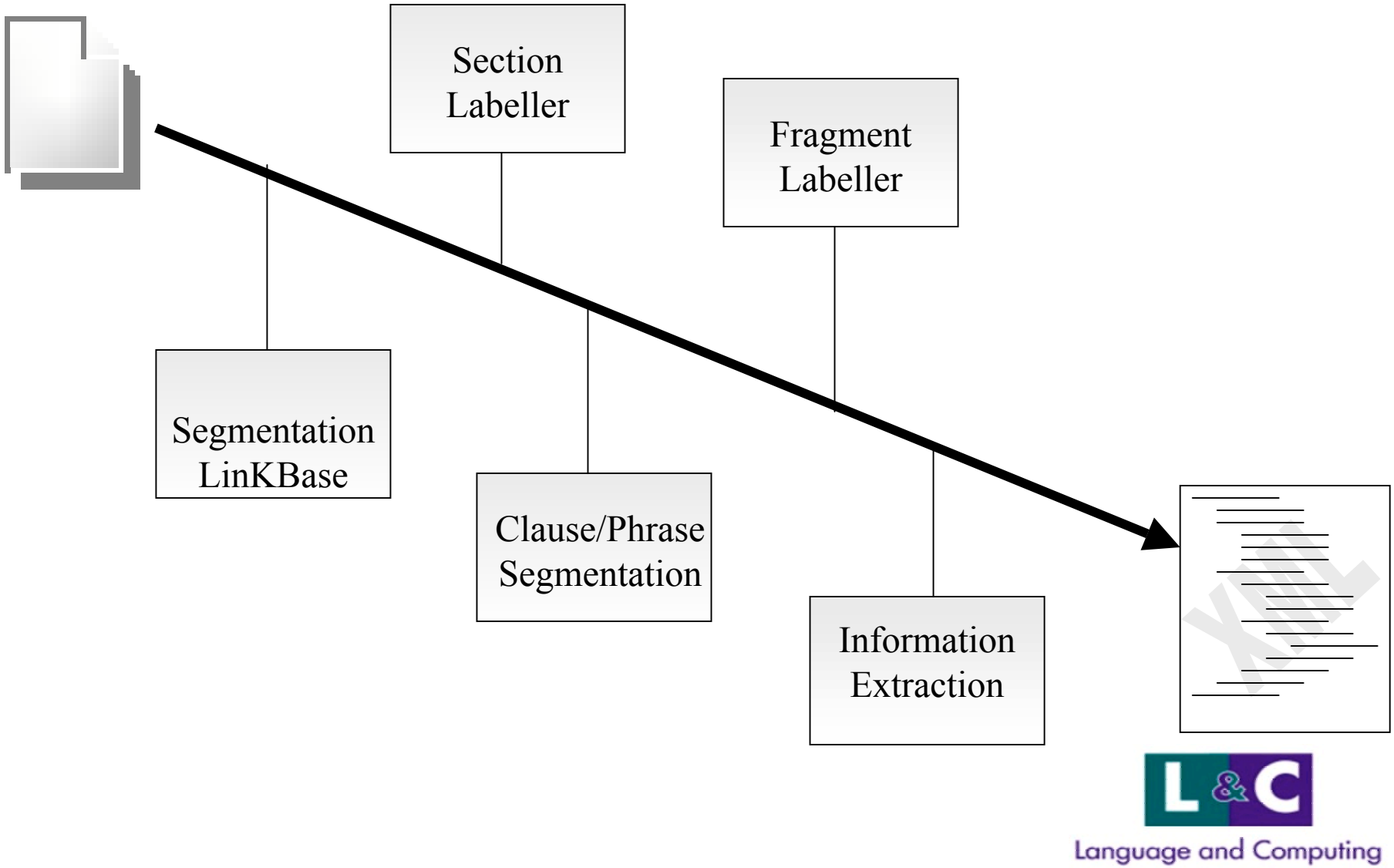
Language and Computing

Medical Information Extraction

- Medical Information Extraction extracts key clinical data from free-text information using a combination of statistical, linguistic and semantic technologies
- Results are stored in XML and can be easily integrated into EMRs, data mining environments (e.g. in pharmaceutical R&D to speed up drug discovery & development), systems to discover adverse drug events (ADEs), etc.

L&C
Language and Computing

Medical Information Extraction Process



L&C
Language and Computing

Information Extraction - results

Reason for admission

Reason for admission
atrial flutter rate controlled with variable block .

Present complaint anamnesis

Present complaint anamnesis
a two day history of generalized weakness and bilateral arm numbness when he assumes an upright posture .

Allergies

The patient is allergic to codeine which causes pruritus .

Main condition for treatment

Atrial flutter with controlled rate

Surgery performed during current hospitalization

Nature of surgery
ablation of his atrial flutter .

Previous surgery for same/related condition

coronary artery b...
PTCA with a stent plac...

Risk Factors

Risk Factor	Current Risk Factor
Disease Risk Factors	
Hypercholesterolemia	Yes
Hypertension	Yes
Coronary Artery Disease	Yes
Other Risk Factors	
Obesity	Yes

Clinical examinations

Examination	Result
Blood Pressure	138/84
Pulse	85
Respirations	20
Temperature	afebrile

Tests and measurements

Heart investigations

Test Conducted	Result
ECG at Rest	Electrocardiogram on admission was atrial flutter with variable atrioventricular block .
ECG at Rest	No acute ST or T wave changes were noted .
ECG at Rest	Rate was 79 beats per minute .
Electrophysiology Study	Electrophysiologic study with ablation of atrial flutter .

Lab results

Hematology

Labaratory test	Result	Normal Range
Hematocrit	39	Male 41 - 50% Female 35 - 46%
Hemoglobin	13.1	Male 13.8 - 17.2 g/dL ...emale 12.0 - 15.6 g/dL

Labaratory test	Result	Normal Range
Elektrolytes		
Sodium	134	135 - 146 mmol/L
Potassium	4.9	3.5 - 5.3 mmol/L
Chloride	104	95 - 108 mmol/L
Bicarbonate	17	21 - 28 mEq/L
Kidney Function		
BUN (Blood Urea Nitrogen)	32	7 - 30 mg/dL
Creatinine	1.6	<=1.2 mg/d...

HTML View of XML-structured, field-defined data

Medication Extraction - results

Data in Table 'LAndCRxMedDose' in 'Northwind' on 'LIBRARIAN\CMS'

LAndCRxMedDoseI	LAndCRxMedID	Route	Qty	QtyUnits	Takes	TakeFre
292	185	per os	1	<NULL>	<NULL>	every d
293	186	<NULL>	<NULL>	<NULL>	<NULL>	daily
294	187	<NULL>	<NULL>	<NULL>	<NULL>	daily
295	188	<NULL>	<NULL>	<NULL>	2	day
296	188	<NULL>	<NULL>	<NULL>	4	day
297	189	per os	2.0	<NULL>	<NULL>	<NULL>
298	189x	per os	1	<NULL>	3	day
299	189	per os	<NULL>	<NULL>	<NULL>	every d
300	189	per os	2.0	<NULL>	<NULL>	every d
301	189	per os	1	<NULL>	<NULL>	every d
302	189	per os	1	<NULL>	<NULL>	every d

Data in Table 'LAndCRxMedComp' in 'Northwind' on 'LIBRARIAN\CMS'

LAndCRxMedComp	LAndCRxMedID	MedName	MedCode	Strength	StrengthUnits	BaseStr
314	188	pepcid	2345	320.0	milligram	<NULL>
315	188		<NULL>	<NULL>	milligram	<NULL>
316	188	ativan	2345	1	milligram	<NULL>
317	189	isordil	2345	40	milligram	<NULL>
318	189	persantine	2345	75	milligram	<NULL>
319	189	norvasc	2345	25.0	milligram	<NULL>
320	189		<NULL>	<NULL>	milligram	<NULL>
321	189	zestril	2345	5	milligram	<NULL>
322	189	prednisone	2345	8	milligram	<NULL>
323	189	elavil	2345	20	milligram	<NULL>
324	190	lipitor	2345	40	milligram	<NULL>
325	190	coumadin	2345	51.25	milligram	<NULL>
326	190		<NULL>	<NULL>	milligram	<NULL>
327	190	isordil	2345	40	milligram	<NULL>
328	190	persantine	2345	75	milligram	<NULL>

LAndCRxMedTakeT	LAndCRxMedDoseI	TakeTimeSpec
1	6	q. a.m. q. p.m.
2	11	with meals
3	18	q. a.m. q. p.m.
4	41	q. a.m. q. p.m.
5	46	with meals
6	53	q. a.m. q. p.m.
7	76	q. a.m. q. p.m.
8	81	with meals
9	93	q. a.m. q. p.m.
10	98	with meals
11	105	q. a.m. q. p.m.
15	163	q. a.m. q. p.m.
16	168	with meals
17	175	q. a.m. q. p.m.
18	198	q. a.m. q. p.m.
19	203	with meals
20	215	q. a.m. q. p.m.
21	220	with meals
22	227	q. a.m. q. p.m.
23	250	q. a.m. q. p.m.
24	255	with meals
29	320	q. a.m. q. p.m.
30	325	with meals
31	332	q. a.m. q. p.m.
12	128	q. a.m. q. p.m.
13	133	with meals
14	140	q. a.m. q. p.m.
25	262	q. a.m. q. p.m.
26	285	q. a.m. q. p.m.
27	290	with meals
28	297	q. a.m. q. p.m.

L&C

Language and Computing

Biochemistry

Labaratory test	Date/Time	Result	Normal Range
Elektrolytes			
Sodium	n/a	134	135 - 146 mmol/L
Potassium	n/a	4.9	3.5 - 5.3 mmol/L
Chloride	n/a	104	95 - 108 mmol/L
Calcium	n/a	9.0	8.5 - 10.3 mg/dL
Phosphorus	n/a	4.0	2.5 - 4.5 mg/dL
Bicarbonate	n/a	17	21 - 28 mEq/L
Lipid Panel			
Kidney Function			
BUN (Blood Urea Nitrogen)	n/a	32	7 - 30 mg/dL
Creatinine	n/a	1.6	<=1.2 mg/dL
Albumin	n/a	3.9	3.5 - 5.0 g/dL
Total proteins	n/a	6.5	6.0 - 8.5 g/dL
Liver Function			
Biluribine total	n/a	0.6	<=1.3 mg/dL
Biluribine direct	n/a	0.1	<0.4 mg/dL
GGT (Gamma Glutamyl Transferase)	n/a	47	Male <=65 U/L Female <=45 U/L
GOT (AST)	n/a	15	<=42 U/L
Albumin	n/a	3.9	3.5 - 5.0 g/dL
Total proteins	n/a	6.5	6.0 - 8.5 g/dL
Diabetes			
Glucose	n/a	133	70 - 125 mg/dL
Heart Function			
CK (Creatine Kinase)	n/a	52	Male <=235 U/L Female <=190 U/L
CK-MB	n/a	fraction	< 3% of total
LDH (Lactate Dehydrogenase)	n/a	217	<=270 U/L
Troponin I	n/a	0.04	0.6 ng/mL

Template

Language and Computing

Accuracy

Patient Number
Patient Name
Letter Type
Dictating Date
Transcription Date
Transcription Time
Dictating Physician Number
Agegroup
PDF Number
Admission Hospital
Patient Visit
Patient Type
Date of Birth
Patient Account
Admission Date
Age (on Admission)
Gender
Discharge Date
Age (on Discharge)
Age (in Text)
Weight
Current Medication
Potassium
Sodium
Allergies

Precision

Better for some parameters than others:

- For the best 25: 100 %
- For the best 30: over 95 %
- For all 55: avg. 87.7 %

Language and Computing

Accuracy

Chest X-ray
Reason for Admission
Previous Surgery Nature
Patient Routing
Smoker
Main Condition for Treatment
Discharge Condition
Weight
Discharge Diagnosis
Discharge To
ECG at Rest
Risk Factor
Catheterization Results
Echocardio at Rest
Present Complaint Anamnes
Hematocrit
Present Surgery Nature
Admission Medication
Pulse
Creatinine
Blood Pressure
Current Medication
Sum
CORRECTED AUTOMATIC E
Diet
Instructions
Potassium
Respirations
Follow-up Physician Name
BUN

Recall

Better for some parameters than others:

- For the best 18: 100 %
- For the best 27: over 90 %
- For all 55: avg. 84.5%

Language and Computing

Accuracy

For medication only :

	Precision	Recall
•Discharge	90.4	96.1
•Current	100	80.8
•Admission	93.9	74.3

Language and Computing

Conclusion

Medical NLP delivers an intelligent layer to the automated processing of clinical data by healthcare IT systems, resulting in :

- Faster healthcare
- Better clinical decisions
- Less medical / medication errors
- Faster availability of more effective

All saving more peoples' lives !

Language and Computing

L&C won the Frost & Sullivan Healthcare Information Technology & Life Sciences "Product of the Year 2003" Award

L&C
Language and Computing

QUESTIONS ?

For further information:

Maarten Laga, CEO

Language and Computing

+32 53 62 95 45

Maarten@landc.be

www.landcglobal.com

L&C

Language and Computing