Extracting Rights and Obligations from Regulations: Towards a Tool-Supported Process

Nadzeya Kiyavitskaya

Nicola Zeni
Luisa Mich
University of Trento
Italy

Travis D. Breaux
Annie I. Antón
College of Engineering
North Carolina State
University, USA

James R. Cordy
Queens University,
Canada
John Mylopoulos
University of Toronto,
Canada

Contents

The Problem

The Manual Methodology

Cerno – Semantic Annotation Framework

Adaptation of Cerno

Preliminary Results

Discussion

The Problem

- ## Alignment challenge

  - Bring policy, law, software requirements and system implementation into agreement

  - Interpret and define system requirements given a policy document

- ## Regulatory documents

  - Legislations are in textual format written in "legalese"

  - Large numbers of internal and external references to other provisions: "as stated in 40 CFR section 262.14(a)(2)"

  - Requirements elaborated at different levels of detail

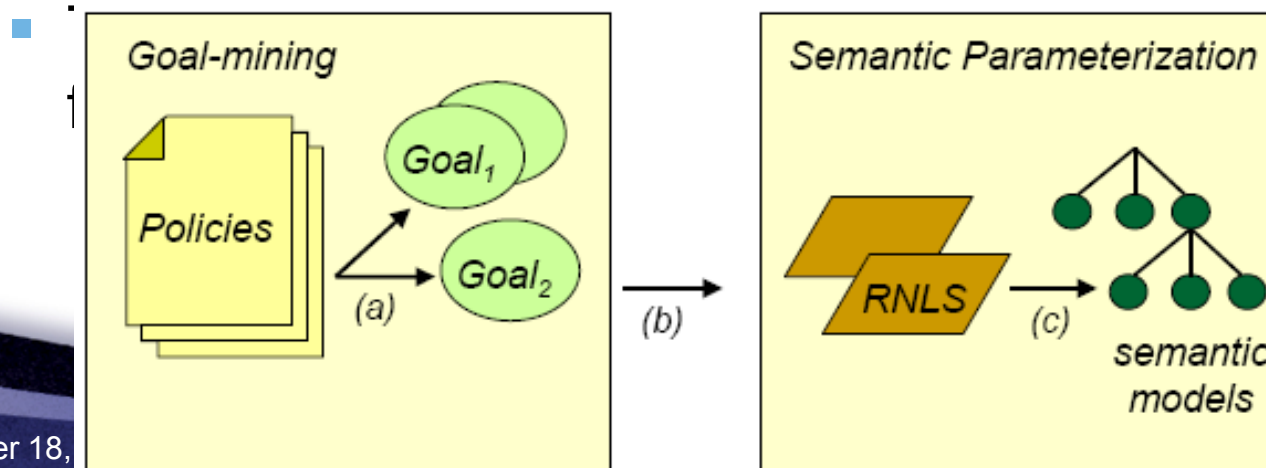 - Statements are prescriptive specifying what must be done

"Extracting Rights and Obligations from Regulations"
University of Toronto

- **The process for extracting requirements from regulations**

  - Regulatory text is annotated to identify text fragments describing actors, rights, obligations, etc.

  - A semantic model is constructed from these annotations

  - ...et of f...

Goal-mining

Policies

Goal₁

Goal₂

(a)

(b)

Semantic Parameterization

RNLS

(c)

semantic models

- **The process for extracting requirements from regulations**

  - Regulatory text is annotated to identify text fragments describing actors, rights, obligations, etc.
  - A semantic model is constructed from these annotations

  Our focus

  - The semantic model is transformed into a set of functional and nonfunctional requirements
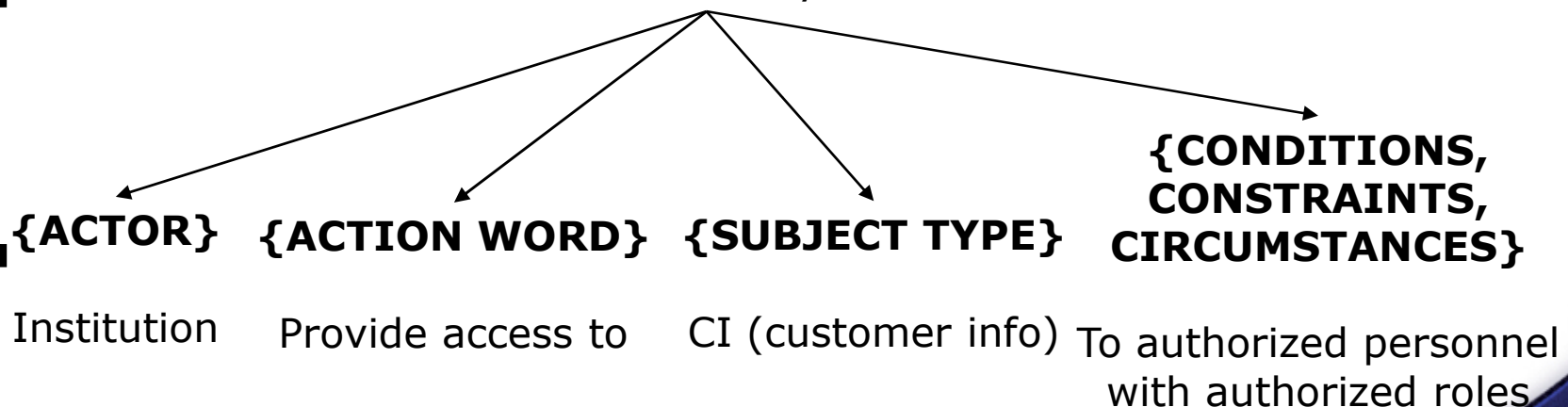
- ## Example: from text to goals in RNLS

Privacy Statement

E.g. Employees are authorized to access customer information only when they need it, to provide you with accounts and services or to maintain your accounts.

Mining Process {

{ACTOR} **{ACTION WORD}** **{SUBJECT TYPE}** **{CONDITIONS, CONSTRAINTS, CIRCUMSTANCES}**

Institution Provide access to CI (customer info) To authorized personnel with authorized roles

Deriving semantic models

RNLS 1: The provider may share information with third parties

```
activity [ right : provider ]
  {
      actor = provider
      action = share
      object = information
      target = third parties
  }
```

RNLS 2: The provider may share information with whom?

```
activity [ right : provider ]
{
      actor = provider
      action = share
      object = information
      target = ? whom
}
```

Previous experience

- Breaux and Antón: developing a theory for understanding rights and obligations
    - Manual markup of the HIPAA Privacy Rule identifying rights or obligations, associated constraints, and conditions
    - Semi-formal rights, obligations and constraints were formally modeled in first-order predicate logic

- Result
    - Phrase heuristics (keyword-based patterns) that indicate rights, obligations, and conditions

The Manual Methodology

Table 1: Patterns for Rights

ID	Pattern	R	A
R_1	<actor> should/may be able to <verb> …	0	0
R_2	<actor> may <verb> …	119	17
R_3	<actor> can/could <verb> …	0	9
R_4	<policy> permits <actor> to <verb> …	3	1
R_5	<actor> would not have to <verb> …	0	0
R_6	<policy> does not restrict… <actor> …	0	0
R_7	<policy> does not require <actor> …	0	0

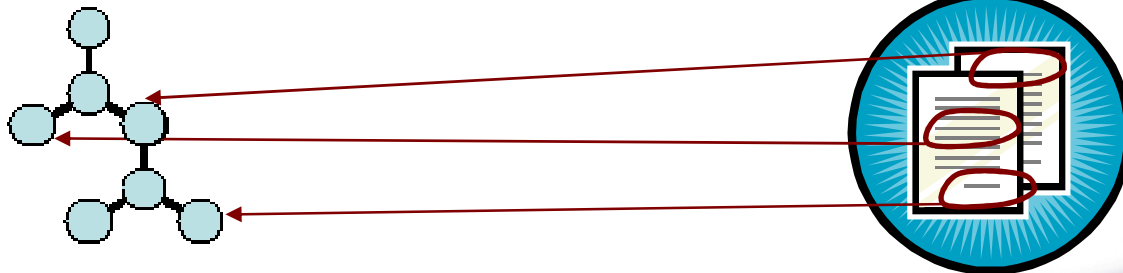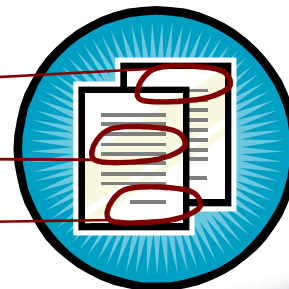- ## **What is semantic annotation?**

  - The annotation of a world-wide web or other natural language document in order to assign explicit real-world meaning to its elements

  - Normally, is based on a semantic model of the domain the document is about

Semantic Model

Documents

"Extracting Rights and Obligations from Regulations"
University of Toronto

- ## The semantic annotation process in Cerno

  - Uses methods and tools proven effective and scalable in the software analysis domain

  - Some successful solutions to the "year 2000" problem utilized design recovery - LS/2000 system
    - analyzed software source code for year 2000 risks and guided a source transformation
    - processed billions of lines of legacy code
    - implemented in **TXL (www.txl.ca)**

  - Factors out domain independent reusable rules and domain specific rules

Cerno – Semantic Annotation Framework

Domain Independent Components

- Object Grammars
- Phrase Grammars

- XML Grammar

- Schema Grammar

Input Documents

Parse

Pre-parsed Documents

Markup

Annotated Documents

Mapping

Database

Domain Dependent Components

- Annotation Schema

- Database Schema Template

"Extracting Rights and Obligations from Regulations"
University of Toronto
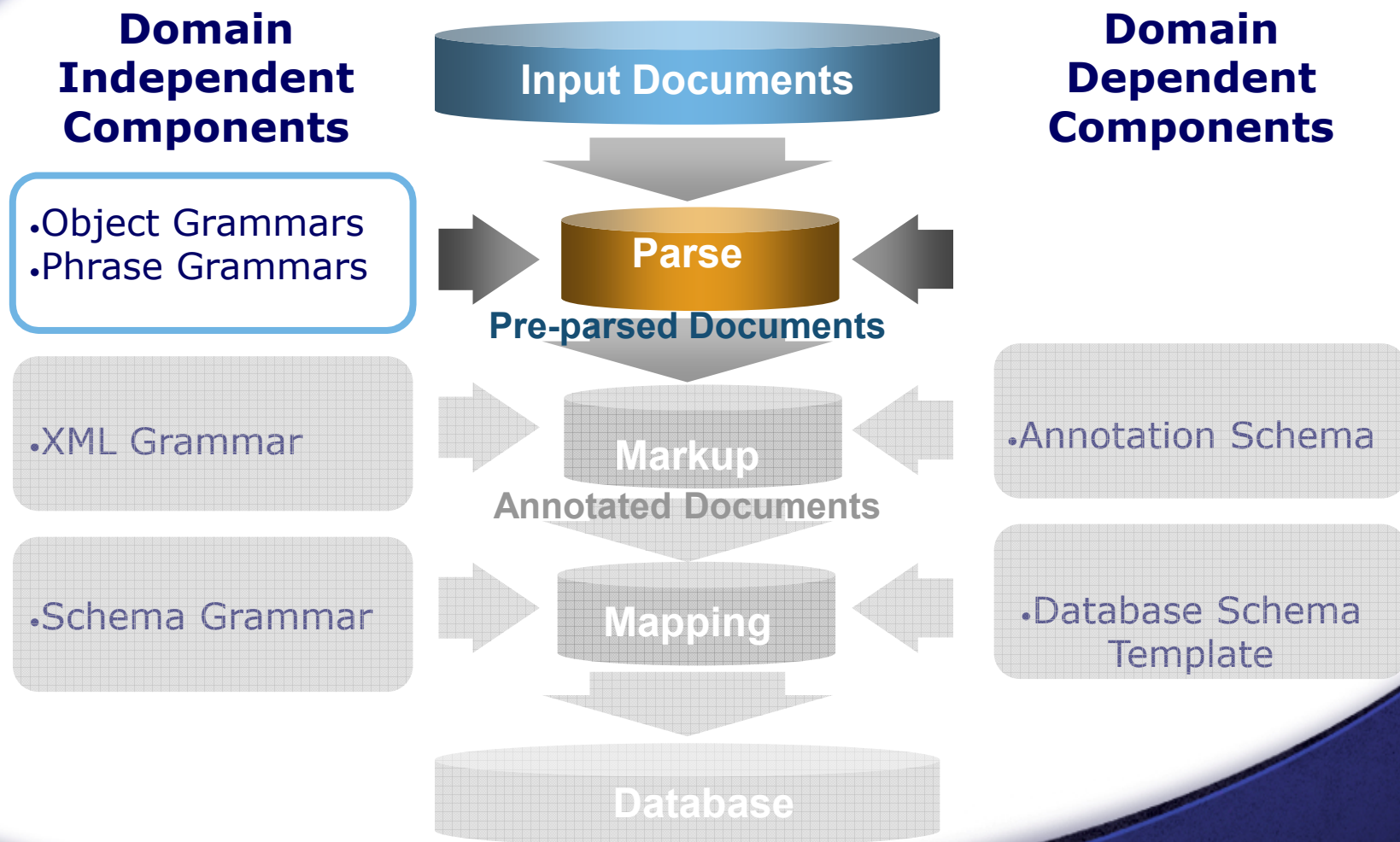
1. Parse

- Non-linguistic context-independent parsing

- Recognizes document elements
  - document, paragraph, sentence, word
  - complex word-equivalent objects
    - `<money>EURO: 115.00/135.00</money>`
    - `<email>alinat@gmail.com</email>`
    - `<phone>+39 3385227285</phone>`
    - `<webaddress>www.residencestandrews.it</webaddress>`
    - `<crossref>§ 164.520 (b)(2)(A)</crossref>`

Cerno – Semantic Annotation Framework

Domain Independent Components

Domain Dependent Components

Input Documents

Parse

Pre-parsed Documents

- Object Grammars
- Phrase Grammars

- XML Grammar

Markup

Annotated Documents

- Annotation Schema

- Schema Grammar

Mapping

- Database Schema Template
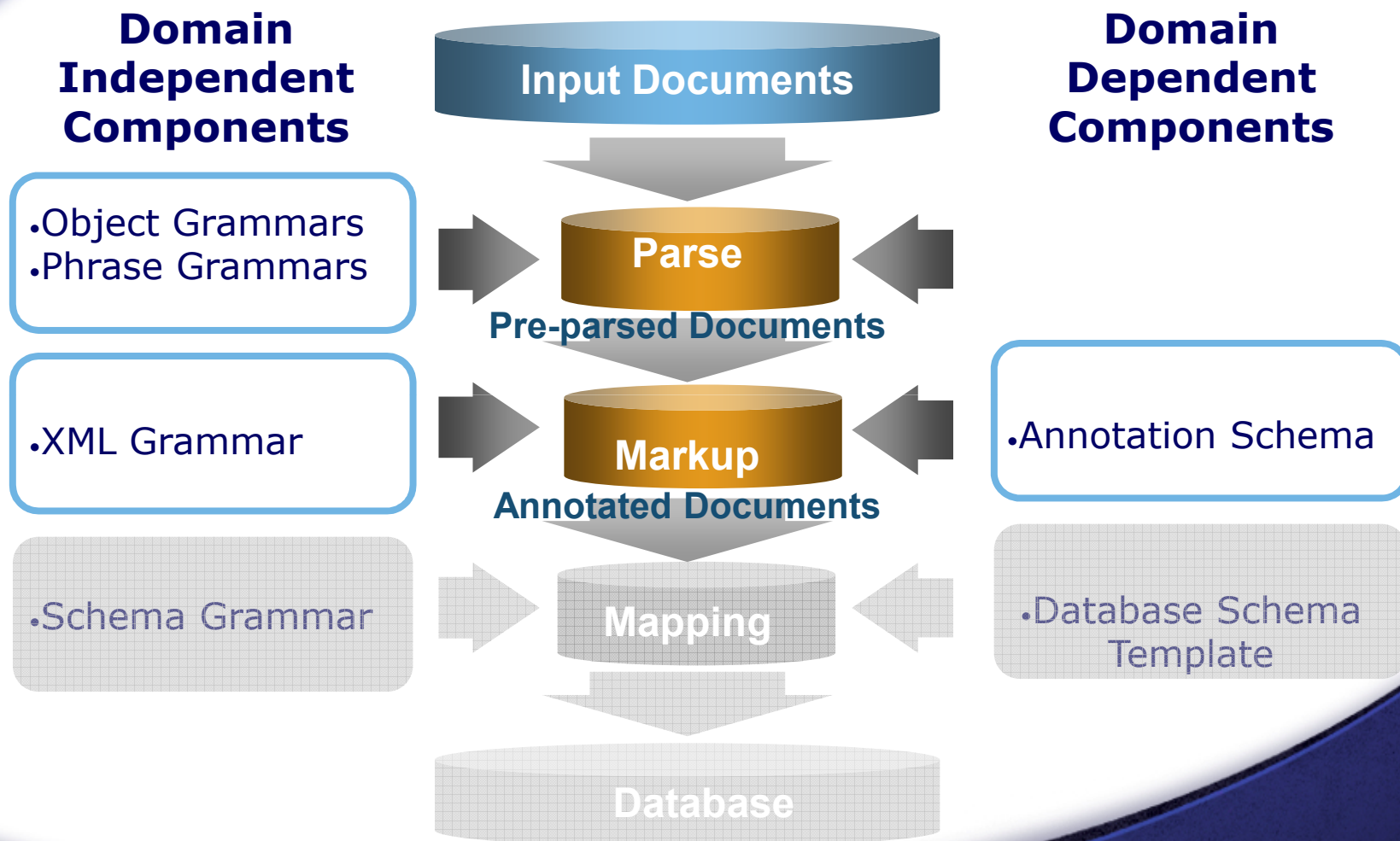
Database

14

"Extracting Rights and Obligations from Regulations"
University of Toronto

2. Markup

- Uses an annotation schema: a wordlist file with indicators for semantic categories, for example
  - Literal word and phrases

    `Location:` `center, located in, near to`

    `Facility:` `air conditioning, air cond.`
  - Names of basic objects

    `Contact:` `phone, email, webaddress`
- Indicators are drawn from the domain semantic model or from examples
- Phrases are marked up once for each category they match

Cerno – Semantic Annotation Framework

Domain Independent Components

Domain Dependent Components

Input Documents

- Object Grammars
- Phrase Grammars

Parse

Pre-parsed Documents

- XML Grammar

Markup

Annotated Documents

- Annotation Schema

- Schema Grammar

Mapping

- Database Schema Template

Database

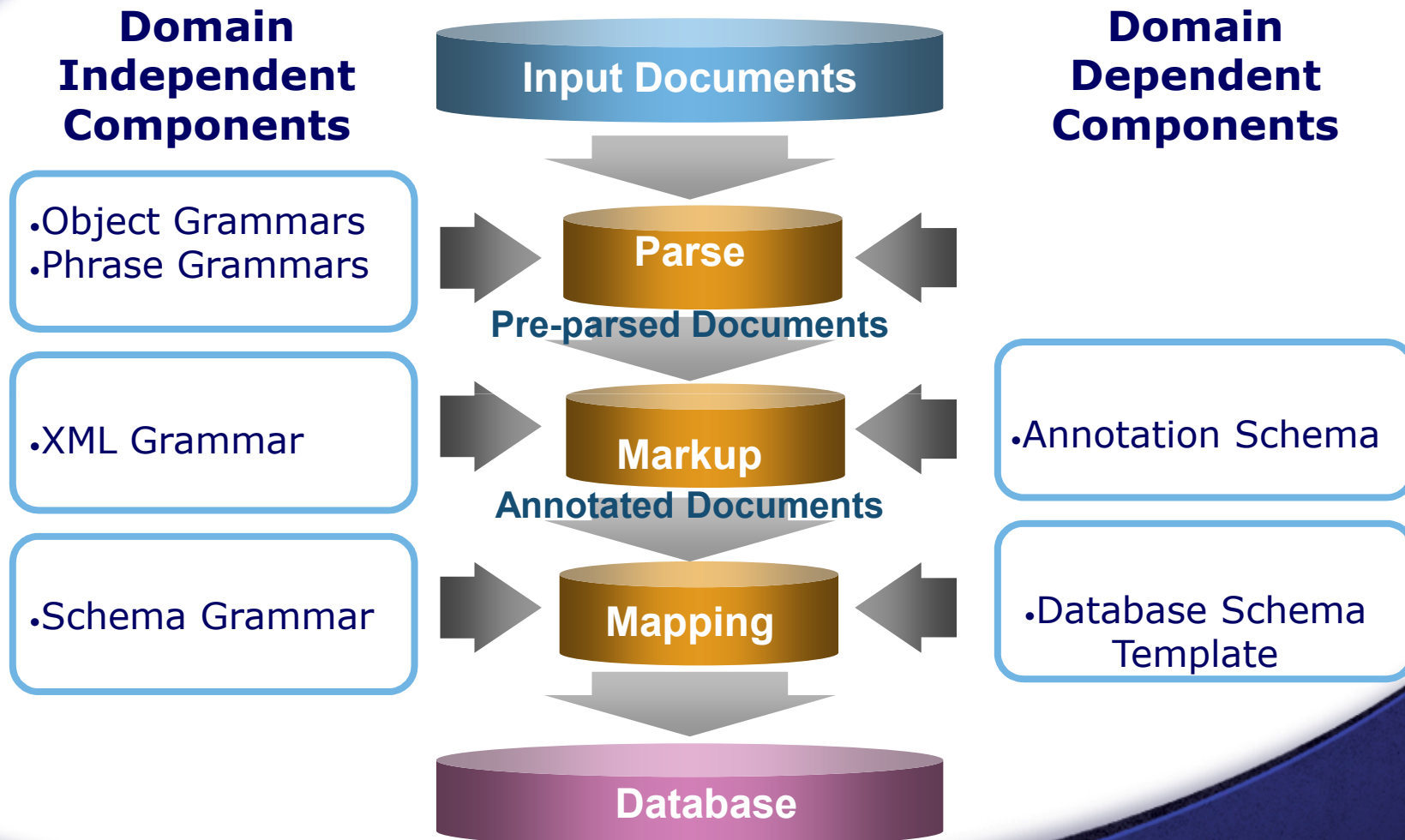"Extracting Rights and Obligations from Regulations"
University of Toronto

3. Mapping

- Populates an XML database schema defined by a user
- Phrases with multiple markups are "cloned"
- The filled database can be then queried by a standard SQL database engine

The framework was tested in two case studies

- Accommodation ads
  - 79% time saving, 0.92 Recall, 0.75 Precision
- Tourist Board Web sites
  - 75% time saving, 0.71 Recall, 0.73 Precision

Cerno – Semantic Annotation Framework

Domain Independent Components

- Object Grammars
- Phrase Grammars

- XML Grammar

- Schema Grammar

Input Documents

→ **Parse** ←

Pre-parsed Documents

→ **Markup** ←

Annotated Documents

→ **Mapping** ←

Database

Domain Dependent Components

- Annotation Schema

- Database Schema Template

"Extracting Rights and Obligations from Regulations"
University of Toronto

Adaptation of Cerno

Goals of the study

- Apply Cerno to the Privacy Rule from HIPAA (the U.S. **H**ealth **I**nsurance **P**ortability and **A**ccountability **A**ct)
  - Provides privacy guidelines to patient and consumers
  - Details conditions for enforcing patient and provider rights, obligations and sanctions

- Support identification of instances of concepts: right, anti-right, obligation, anti-obligation, exception, some types of constraints

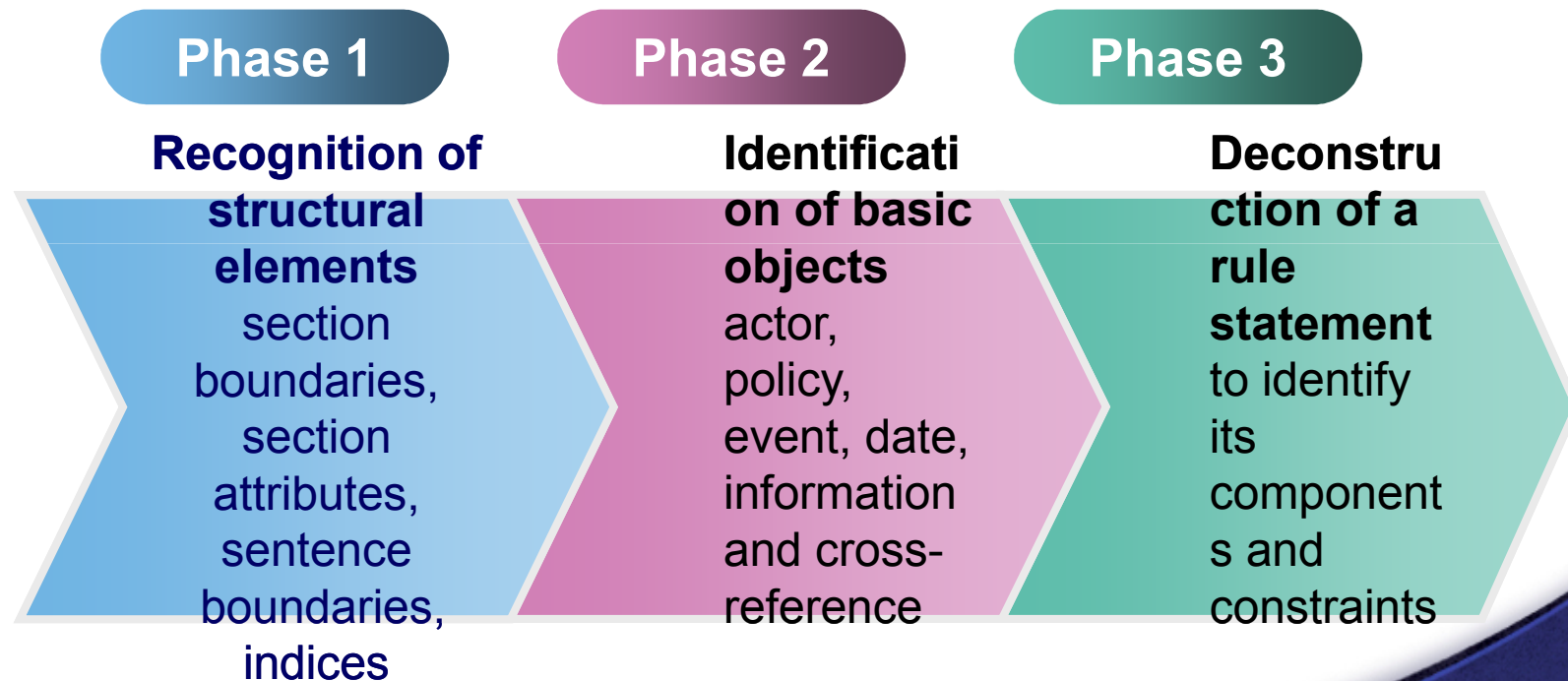- Use the manually identified phrase heuristics

Adaptation of Cerno

- ## **Concepts of interest**

  - Right: an action that a stakeholder is conditionally permitted to perform

  - Obligation: an action that a stakeholder is conditionally required to perform

  - Anti-right: are (implied) obligations for stakeholders
    - The provider may not share information → must not share

  - Anti-obligation: are (implied) rights for stakeholders
    - The provider is not required to cooperate → may cooperate

  - Constraint phrase: the part of a rule statement that describes a single pre-condition

"Extracting Rights and Obligations from Regulations"
University of Toronto

Adaptation of Cerno

- ## The Cerno-based regulation analysis process

Phase 1	Phase 2	Phase 3
Recognition of structural elements section boundaries, section attributes, sentence boundaries, indices	**Identification of basic objects** actor, policy, event, date, information and cross-reference	**Deconstruction of a rule statement** to identify its components and constraints

Adaptation of Cerno

- **Phase 1: Identification of structural elements**
  - Section boundaries, section attributes (number and title), sentence boundaries using user-defined grammar rules

```
define program
    [repeat section]
end define
define section
    [title]
    [repeat sentence]
end define
define title
    `§ [number]  [repeat token_not_dot] [fullstop]
end define
```

```
define sentence
    [repeat phrase_comma] [fullstop]
end define
define fullstop
    `.
end define
```

Phase 2: Basic entities recognition

- Hyponyms derived from section 160.103 "Definitions of HIPAA"

```
Actor: ANSI, business associate(s), covered entit(y
   | ies), HCFA, HHS, health care provider(s), <...>;
Policy: act(s), implementation specification(s),
   standard(s), trading partner agreement(s), covered
   function(s), organized health care arrangement(s);
Information: individually identifiable health
   information, protected health information, health
   information, designated record set(s),
   psychotherapy notes;
```

Adaptation of Cerno

Phase 3: Rule statement analysis

- Uses manually identified normative phrases in HIPAA

Right

```
<actor>...</actor> may

<actor>...</actor> can

<actor>...</actor> could

<policy>...</policy> permits

<actor>...</actor> has a right to

<actor>...</actor> should be able to
```

Temporal Constraint

```
for the time
during
no later
within the time
```

"Extracting Rights and Obligations from Regulations"
University of Toronto

Adaptation of Cerno

- Result annotation

<?xml version="1.0" encoding="ISO-8859-1"?>
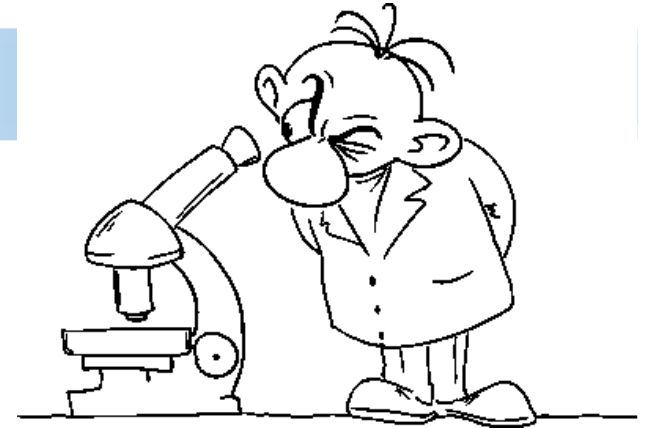
<DOC><Section><Num_Section>§164.528</Num_Section><Title_section> Accounting of <Event>disclosures</Event> of <Information>protected health information</Information></Title_section>. <…>

<Sentence><Index>(ii)</Index> If the agency or official statement in <CrossRef>paragraph (a)(2)(i) of this section</CrossRef> is made orally,<Obligation> the <Actor>covered entity</Actor> must:</Obligation>

<Index>(A)</Index><Continue_Obligation> Document the statement, including the identity of the agency or official making the statement;</Continue_Obligation>

<Index>(B)</Index><Continue_Obligation> Temporarily suspend the <Actor>individual</Actor>'s right to an accounting of <Event>disclosures</Event> subject to the statement; and</Continue_Obligation>

<Index>(C)</Index><Continue_Obligation> Limit the temporary suspension to no longer than 30 days from the date of the oral statement,<Exception> unless a written statement *<Constraint>pursuant to <CrossRef>paragraph (a)(2)(i) of this section</CrossRef></Constraint>* is submitted during that time</Exception></Continue_Obligation></Sentence>.

<…>

Experimental evaluation

- ## Goal

  - Test the usefulness of the tool for non-experts in the regulatory text who need to analyze such documents to generate requirements for a new software system

- ## Settings

  - 4 junior researchers from the software engineering area

  - 2 parts of section 164.520 of HIPAA given for annotation (a total of 2269 words)

    - With tool
    - Without tool

Experimental evaluation

- Quantitative impact of the tool
  - 10% increase in the total number of entities identified
  - annotators were faster by about 12.3 %

- Qualitative impact of the tool
  - All annotators expressed satisfaction with annotations provided by the tool
  - Cerno identified the concepts highly relevant for a requirements engineer

"Extracting Rights and Obligations from Regulations"
University of Toronto

Discussion

- ## Open issues
  - Additional types of constraints should be considered
  - Correct identification of a subject to which a constraint applies
  - Identification of the subjects of conjunctions or disjunctions ("or", "and")
  - Flattening the regulation text

- ## Promises
  - Good quality results with limited effort
  - Unlike human annotations, automatic annotations are more consistent
  - Essential time savings as technology evolves

- ## **Accepted publication**
    - N. Kiyavitskaya, N. Zeni, T.D. Breaux, A.I. Antón, J.R. Cordy, L. Mich, J. Mylopoulos: "*Extracting Rights and Obligations from Regulations: Toward a Tool-Supported Process*", to be published in proc. of ASE '07 (The International Conference on Automated Software Engineering) Atlanta, Giorgia, USA

Thank You !

Feedback and questions?

nadzeya@dit.unitn.it