



# **Getting Grey Literature Out from the Shadow of Inaccessibility**

**Elizabeth Liddy  
Jana Bradley  
Anne Turner**

**March 28, 2001**

---

**Center for NLP**



# Goal

---

- **To provide Public Health (PH) professionals and policy makers with improved access to the grey literature reporting on Public Health interventions**



# Goal

---

- **To provide Public Health (PH) professionals and policy makers with improved access to the grey literature reporting on Public Health interventions**
- **By utilizing Natural Language Processing to provide a universally accessible web site for:**
  - searching
  - summarization
  - navigation
  - visualization



# Definitions

---

- **Public Health interventions - community based programs that treat, prevent, or educate about diseases or health risks**



# Definitions

---

- **Public Health interventions - community based programs that treat, prevent, or educate about diseases or health risks**
- **Grey Literature – information produced by government, academia, or business - but not controlled by commercial publishers**

# Definitions

---

- **Public Health interventions - community based programs that treat, prevent, or educate about diseases or health risks**
- **Grey Literature – information produced by government, academia, or business - but not controlled by commercial publishers**
  - Non-conventional; fugitive; ephemeral
  - 30% of Public Health grey literature is available on the web



# Problem

---

- **Many PH interventions are conducted**
- **Reported in local or state reports, which are neither indexed, catalogued, regularly collected, nor widely distributed**
- **Those reports of interventions on the web are:**
  - Very lengthy, multi-faceted
  - Broad scoped
  - Poorly described



# Solution – Provide Improved Access

---

- **Develop a model of PH interventions based on predictable:**
  - participants      - elements
  - processes        - sequencing of processes



# Solution – Provide Improved Access

---

- **Develop a model of PH interventions based on predictable:**
  - participants      - elements
  - processes          - sequencing of processes
- **Adapt our NLP-based information access system to the PH domain**



# Solution – Provide Improved Access

---

- **Develop a model of PH interventions based on predictable:**
  - participants      - elements
  - processes          - sequencing of processes
- **Adapt our NLP-based information access system to the PH domain**
- **Extract & make searchable the necessary information for determining the potential usefulness of reports**



# Draft Model of a PH Intervention

---

- PH Problem
- Health Professionals
- Sponsoring Agency
- Intervention Type
- Target Population
- Setting
- Outcomes
- Evaluation
- Recommendations
- Limits of Study



# Natural Language Processing

---



# Natural Language Processing

---

- a range of computational techniques



# Natural Language Processing

---

- a range of computational techniques
- for analyzing and representing naturally occurring texts

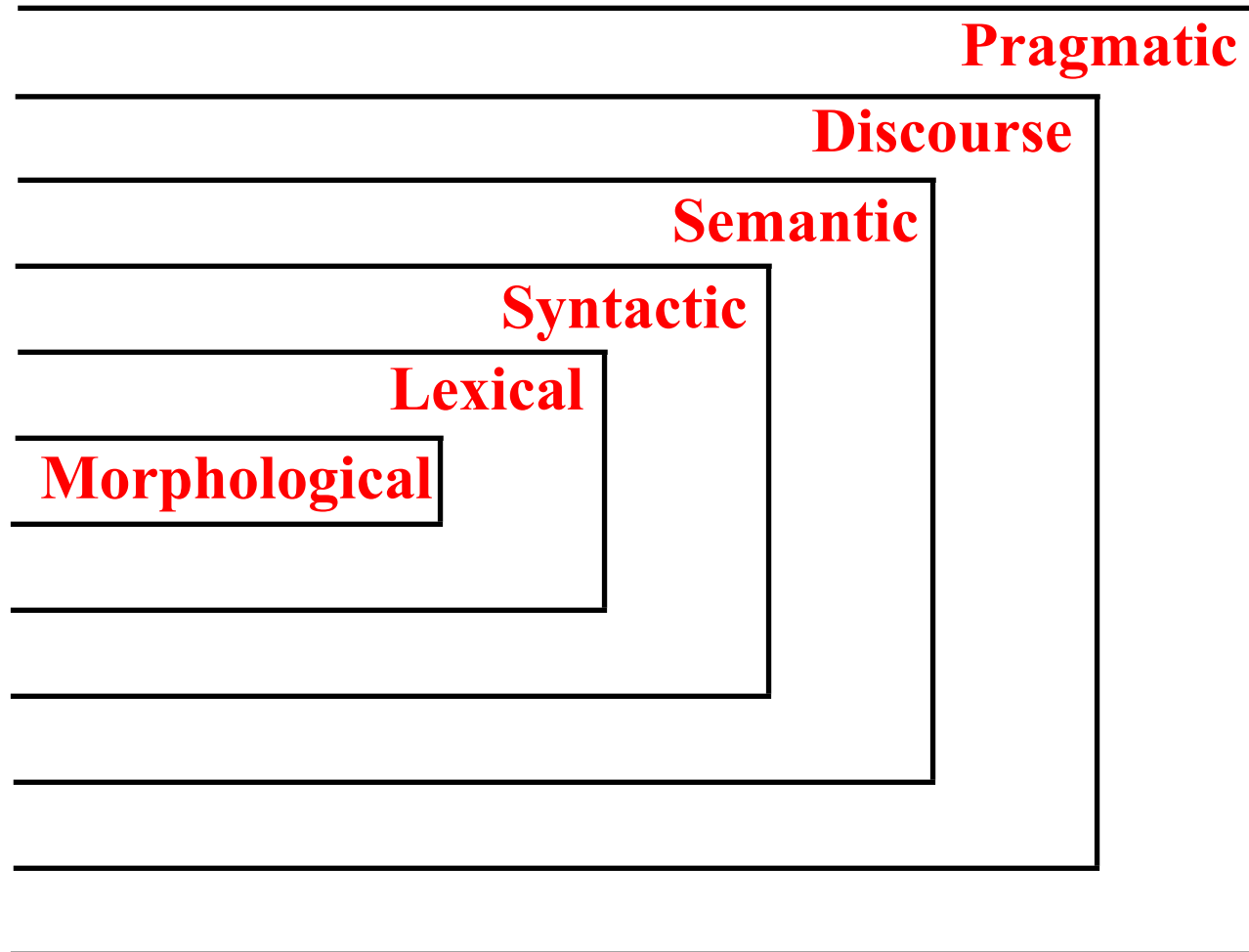


# Natural Language Processing

---

- a range of computational techniques
- for analyzing and representing naturally occurring texts
- at one or more levels of linguistic analysis

# Levels of Language Understanding





# Natural Language Processing

---

- a range of computational techniques
- for analyzing and representing naturally occurring texts
- at one or more levels of linguistic analysis
- for the purpose of achieving human-like language processing

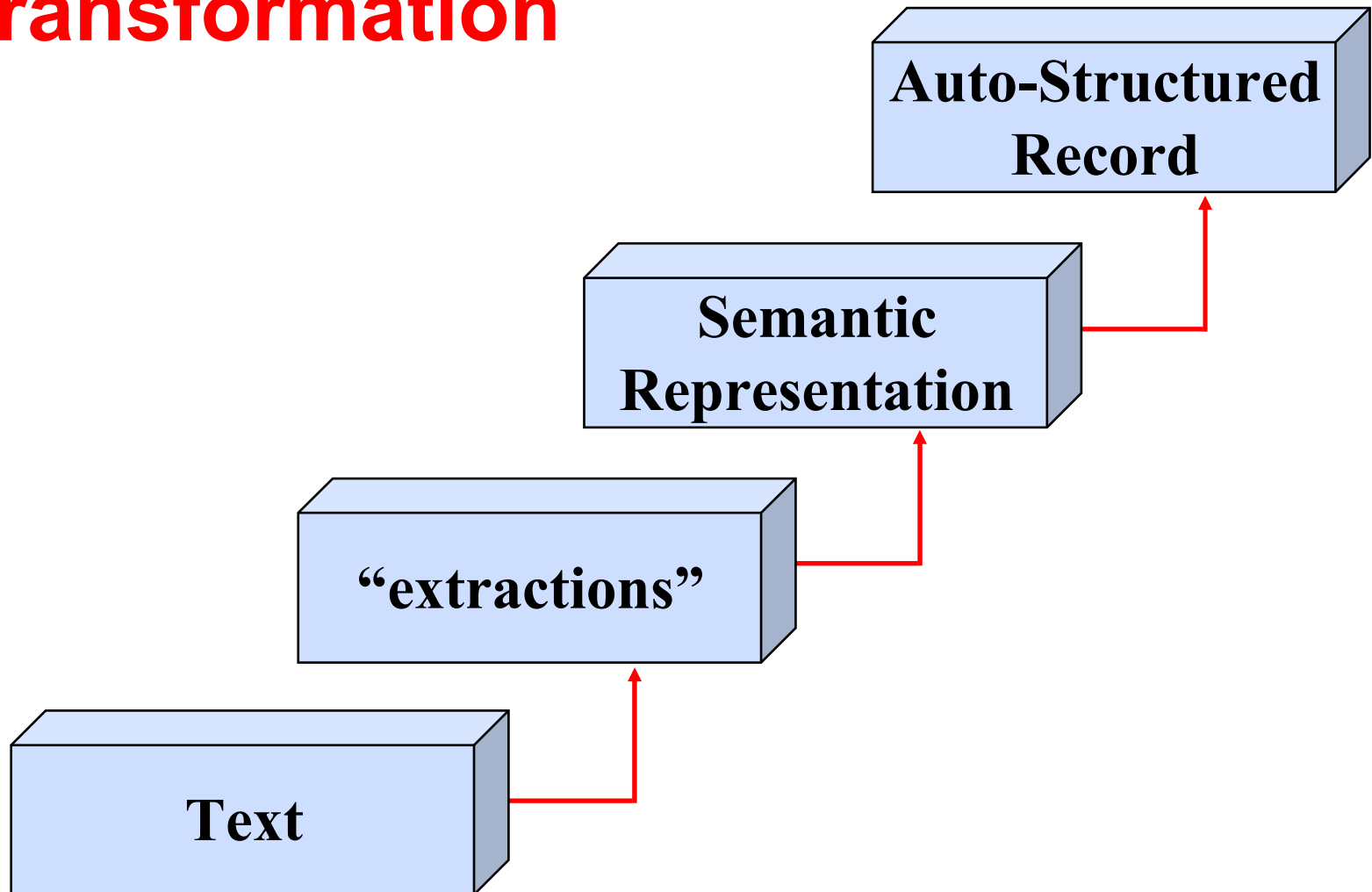


# Natural Language Processing

---

- a range of computational techniques
- for analyzing and representing naturally occurring texts
- at one or more levels of linguistic analysis
- for the purpose of achieving human-like language processing
- for a wide range of applications

# Ladder of Transformation



---

## **Tagging & Extraction Process**

---

**Westco Bancorp has been merged into MAF Bancorp, and First Federal Savings and Loan Association of Westchester, a wholly-owned subsidiary of Westco Bancorp .....**

## Tagging & Extraction Process

Westco Bancorp has been merged into MAF Bancorp, and First Federal Savings and Loan Association of Westchester, a wholly-owned subsidiary of Westco Bancorp .....

Westco|NP Bancorp|NP has|VBZ been|VBN merged|VBN into|IN MAF|NP Bancorp|NP ,|, and|CC First|NP Federal|NP Savings|NP and|CC Loan|NP Association|NP of|IN Westchester|NP ,|, a|DT wholly-owned|VBN subsidiary|NN of|IN Westco|NP Bancorp|NP.....

## Tagging & Extraction Process

**Westco Bancorp has been merged into MAF Bancorp, and First Federal Savings and Loan Association of Westchester, a wholly-owned subsidiary of Westco Bancorp .....**

**Westco|NP Bancorp|NP has|VBZ been|VBN merged|VBN into|IN MAF|NP Bancorp|NP ,|, and|CC First|NP Federal|NP Savings|NP and|CC Loan|NP Association|NP of|IN Westchester|NP ,|, a|DT wholly-owned|VBN subsidiary|NN of|IN Westco|NP Bancorp|NP**

**<NP> Westco\_Bancorp </NP> has|VBZ been|VBN merged|VBN into|IN <NP> MAF\_Bancorp </NP> ,|, and|CC <NP> First\_Federal\_Savings\_and\_Loan\_Association\_of\_Westchester </NP> ,|, a|DT <NP> wholly-owned\_subsidary </NP> of|IN <NP> Westco\_Bancorp </NP>**

## Tagging & Extraction Process

**Westco Bancorp has been merged into MAF Bancorp, and First Federal Savings and Loan Association of Westchester, a wholly-owned subsidiary of Westco Bancorp .....**

**Westco|NP Bancorp|NP has|VBZ been|VBN merged|VBN into|IN MAF|NP Bancorp|NP ,|, and|CC First|NP Federal|NP Savings|NP and|CC Loan|NP Association|NP of|IN Westchester|NP ,|, a|DT wholly-owned|VBN subsidiary|NN of|IN Westco|NP Bancorp|NP**

**<NP> Westco\_Bancorp </NP> has|VBZ been|VBN merged|VBN into|IN <NP> MAF\_Bancorp </NP> ,|, and|CC <NP> First\_Federal\_Savings\_and\_Loan\_Association\_of\_Westchester </NP> ,|, a|DT <NP> wholly-owned\_subsidary </NP> of|IN <NP> Westco\_Bancorp </NP>**

**<NP;type=company;sf=Westco Bancorp Inc> Westco\_Bancorp </NP>**  
**has|VBZ been|VBN merged|VBN into|IN <NP;type=company; sf=MAF**  
**Bancorp Inc> MAF\_Bancorp </NP> ,|, and|CC <NP;type=company; sf=First**  
**Federal Savings and Loan Association of Westchester> First\_Federal\_**  
**Savings\_and\_Loan\_Association\_of\_Westchester </NP> ,|, a|DT <NP> wholly-**  
**owned\_subsidary </NP> of|IN <NP;type=company;sf=Westco Bancorp Inc>**  
**Westco\_Bancorp </NP>**

<NP;type=company;sf=Westco Bancorp Inc> Westco\_Bancorp </NP>  
has|VBZ been|VBN merged|VBN into|IN <NP;type=company; sf=MAF  
Bancorp Inc> MAF\_Bancorp </NP> ,|, and|CC <NP;type=company; sf=First  
Federal Savings and Loan Association of Westchester> First\_Federal\_  
Savings\_and\_Loan\_Association\_of\_Westchester </NP> ,|, a|DT <NP> wholly-  
owned\_subsidary </NP> of|IN <NP;type=company;sf=Westco Bancorp Inc>  
Westco\_Bancorp </NP>

**gen-relation: isa**

**entity1: First Federal Savings & Loan Assoc. of Westchester (company)**

**entity2: subsidiary**

**gen-relation: characteristic**

**entity1: subsidiary**

**entity2: wholly-owned**

**gen-relation: possess**

**entity1: Westco Bancorp Inc (company)**

**entity2: subsidiary**



# Uses of this rich extraction

---

- 1. Information Retrieval**
- 2. Question – Answering**
- 3. Information Extraction**
- 4. Summarization**
- 5. Cross-language Retrieval**
- 6. MetaTagging for Digital Libraries**
- 7. Link Analysis**
- 8. Human-Computer Dialogue**
- 9. Visualization**







# Public Health System Process

---

- **Full-text PH reports are analyzed by Natural Language Processing**



# Public Health System Process

---

- **Full-text PH reports are analyzed by Natural Language Processing**
- **The NLP technology identifies the information to fill the slot for each component in the model**



# Public Health System Process

---

- **Full-text PH reports are analyzed by Natural Language Processing**
- **The NLP technology identifies the information to fill the slot for each component in the model**
- **System extracts & tags this information**



# Public Health System Process

---

- **Full-text PH reports are analyzed by Natural Language Processing**
- **The NLP technology identifies the information to fill the slot for each component in the model**
- **System extracts & tags this information**
- **Produces a structured summary of each report for searching**

<b>Problem</b>	<b>teenage smoking</b>
<b>Health care professionals</b>	<b>public health researchers</b>
<b>Sponsoring agency</b>	<b>Robert Wood Johnson Foundation</b>
<b>Intervention type</b>	<b>information provision</b>
<b>Target population</b>	<b>teenagers</b>
<b>Age</b>	<b>12 to 15 years old</b>
<b>Gender</b>	<b>male and female</b>
<b>Ethnicity</b>	<b>all</b>
<b>Geographic location</b>	<b>Massachusetts</b>
<b>Socioeconomic status</b>	<b>all</b>
<b>Focal unit</b>	<b>individual</b>
<b>Setting</b>	<b>media campaign</b>
<b>Time frame</b>	<b>4 years</b>
<b>Outcomes</b>	<b>Teenagers exposed to a counter-advertising campaign on television were half as likely to become smokers as though who were not.</b>

.....cont'd



# Outcomes

---

- **Enable Public Health personnel to make more informed decisions on best interventions for their communities**



# Outcomes

---

- **Enable Public Health personnel to make more informed decisions on best interventions for their communities**
- **Facilitate creation of rational Public Health policies based on a clear, complete overview of the current state of PH interventions**



# Outcomes

---

- **Enable Public Health personnel to make more informed decisions on best interventions for their communities**
- **Facilitate creation of rational Public Health policies based on a clear, complete overview of the current state of PH interventions via**
  - **Data-mining of records**



# Outcomes

---

- **Enable Public Health personnel to make more informed decisions on best interventions for their communities**
- **Facilitate creation of rational Public Health policies based on a clear, complete overview of the current state of PH interventions via**
  - Data-mining of records
  - Visualizing the shape of the PH domain
    - Reveal areas of need
    - Prevent ‘re-invention of the wheel’
    - Identify best practices