Access Control Policy Extraction from Natural Language Text

John Slankas and Laurie Williams

Motivation

Access control is a paramount concern in health information systems:

- Ensure only authorized users can view and alter confidential records
- Prevent unauthorized users from making malicious changes
- Ensure user trust in the system's integrity
- Compliance with various laws and regulations

Access control remains a significant issue:

- 30% of the CWE SANS Top 25 Most Dangerous Software Issues
- 61% of the incidents in the 2013 Verizon Data Breach Investigations Report include some form of access control abuse

Large number of requirement sources:

- Federal laws and regulations (HIPAA, Certified HER, etc.)
- State laws and regulations
- Industry Guidelines
- Organizational policies and procedures
- Project specific requirements

Goal

Help developers improve security by extracting the access control policies implicitly and explicitly defined in natural language artifacts.

Approach

Apply a combination of natural language processing (NLP), machine learning (ML), and information extraction (IE) techniques to create an interactive process to parse existing, unconstrained natural language texts to extract access control policies

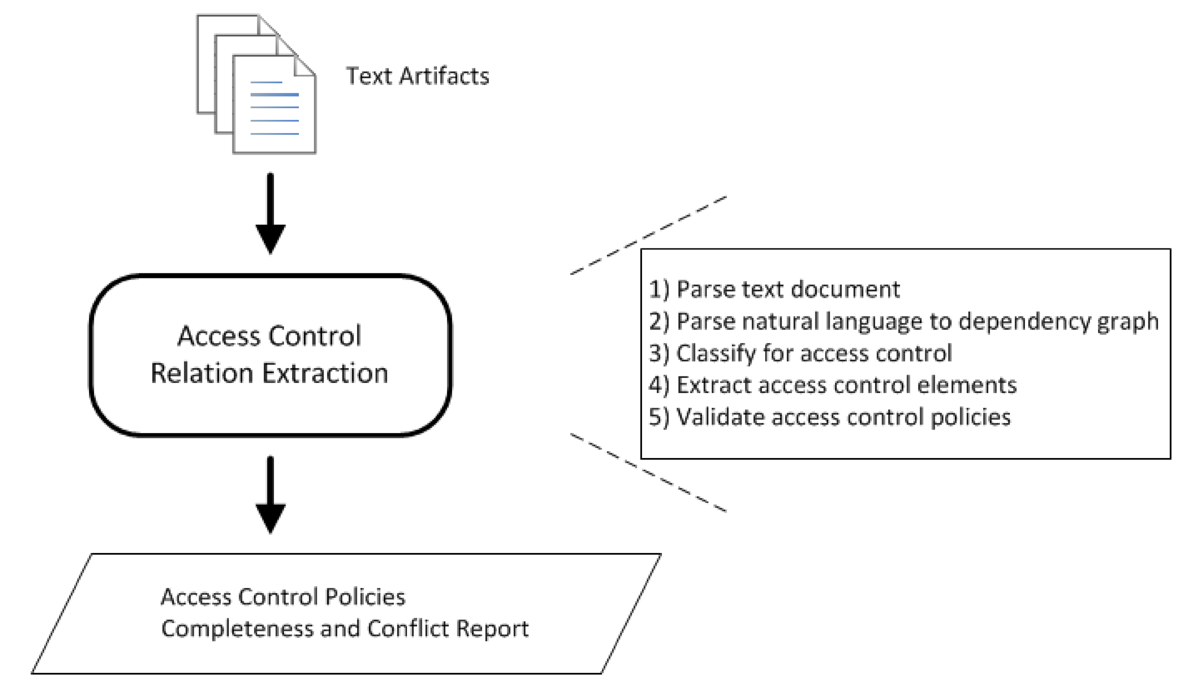
- NLP: Convert the sentence into a dependency graph
- ML: Determine whether or not the sentence contains an access control policy
- IE: Extract access policy using the relations among words to find subjects, actions, and resources

Once extracted, the access control policies can examined for coverage in terms of the subjects versus resources as well as for conflicts. Developers can implement extracted policies or verify existing policies.

Research Questions

1. How effectively can we identify access control policies in natural language text in terms of precision and recall?
2. What common patterns exist in sentences expressing access control policies?
3. What is an appropriate set of seeded graphs to effectively bootstrap the process to extract the access control elements?

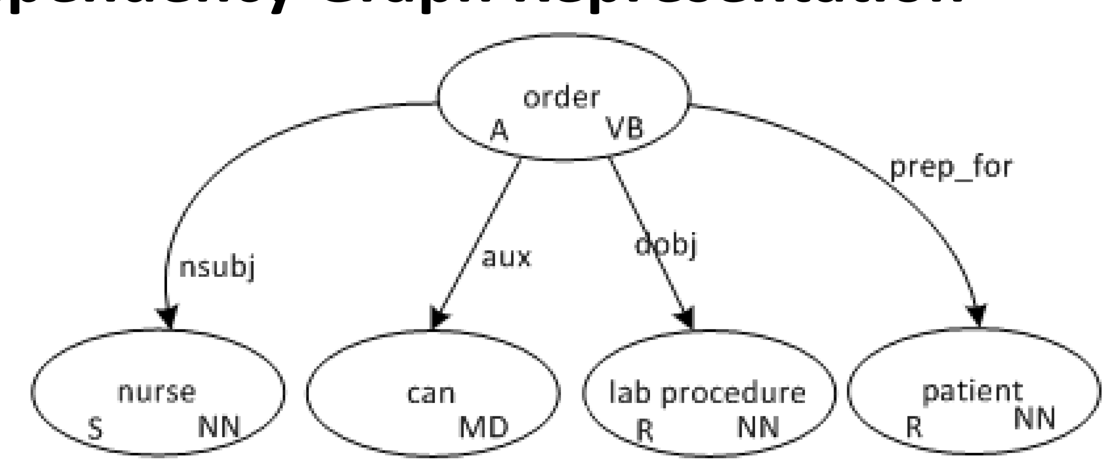
Process Overview



Document Parsing

document → *line*
line → *listID* title *line* | title *line* | *sentence line* | λ
sentence → normalSentence | listStart (":" | "-") *listElement*
listElement → *listID sentence listElement* | λ
listID → *listParanID* | *listDotID* | number
listParanID → "(" *id* ")" *listParanID* | *id* ")" *listParanID* | λ
listDotID → *id* "." *listDotID* | λ
id → letter | romanNumeral | number

Dependency Graph Representation



A nurse can order a lab procedure for a patient.

Access Control Representation

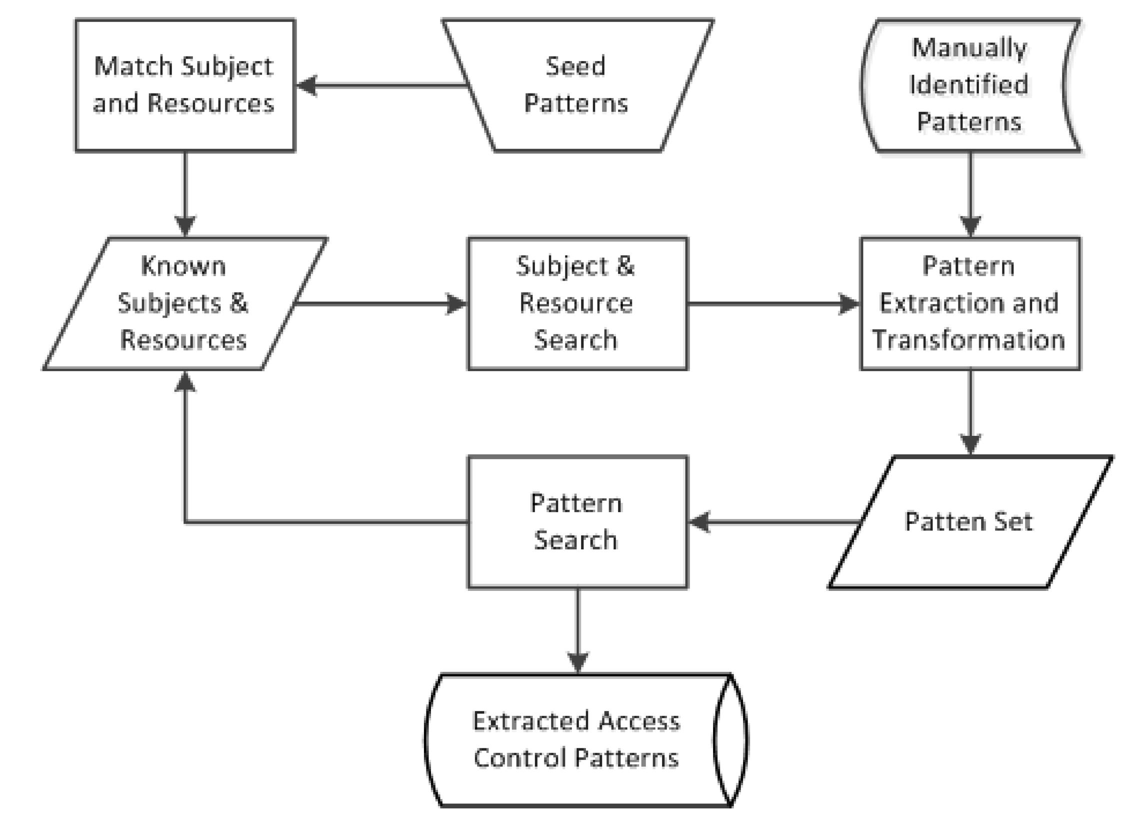
$A(\{s\},\{a\},\{r\},[n],[l],\{c\},H,p)$

s	vertices composing the subject
a	vertices composing the action
r	vertices composing the resource
n	vertex representing negativity
l	vertex representing limitation to a specific role
c	vertices providing context to the access control policy
H	subgraph required to connect all previous vertices
p	set of permission associated with the action

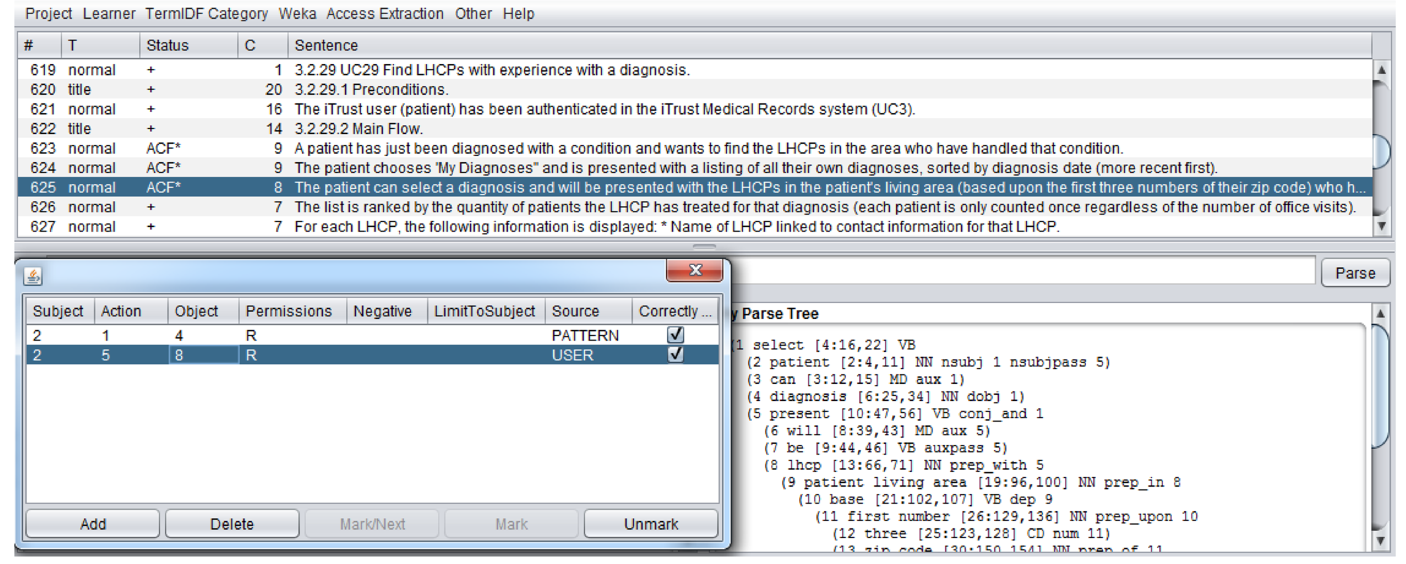
$A((nurse),(order),(lab\ procedure),(\,),(\,),V{:}\,nurse,order,lab\ procedure;$
$E{:}\,(order,nurse,nsubj);(order,lab\ procedure,dobj)\,),create)$

$A((nurse),(order),(patient),(\,),(\,),(V{:}\,nurse,order,patient;$
$E{:}\,(order,nurse,nsubj);(order,patient,prep\_for)\,),read)$

Access Control Extraction



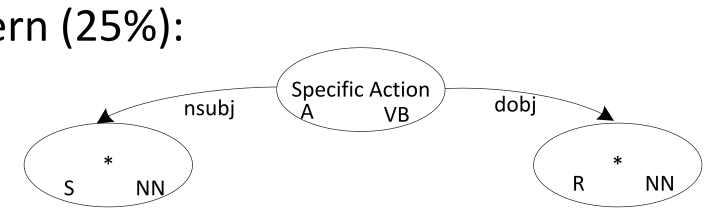
Application Screenshot



Initial Results

Stratified Ten-Fold Cross Validation			
Classifier	**Precision**	**Recall**	**F_1 Measure**
Naïve Bayes	.743	.940	.830
SMO	.845	.830	.837
TF-IDF	.588	.995	.739
k-NN (k=1)	.851	.830	.840
Combined SL	.873	.908	.890

Most common pattern (25%):



Seeding Performance:

With a set of 10 action verbs defined, the process found access control policies with with a precision of .463 and a recall of .536.

Future Work

- Look for addition seeds and ways to improve performance
- Measure impact of user involvement to identify additional patterns
- Refine resource mapping (objects, relations, and attributes)
- Use additional contextual information to limit / control access

North Carolina State University
Department of Computer Science

John.Slankas@ncsu.edu
Laurie_Williams@ncsu.edu