

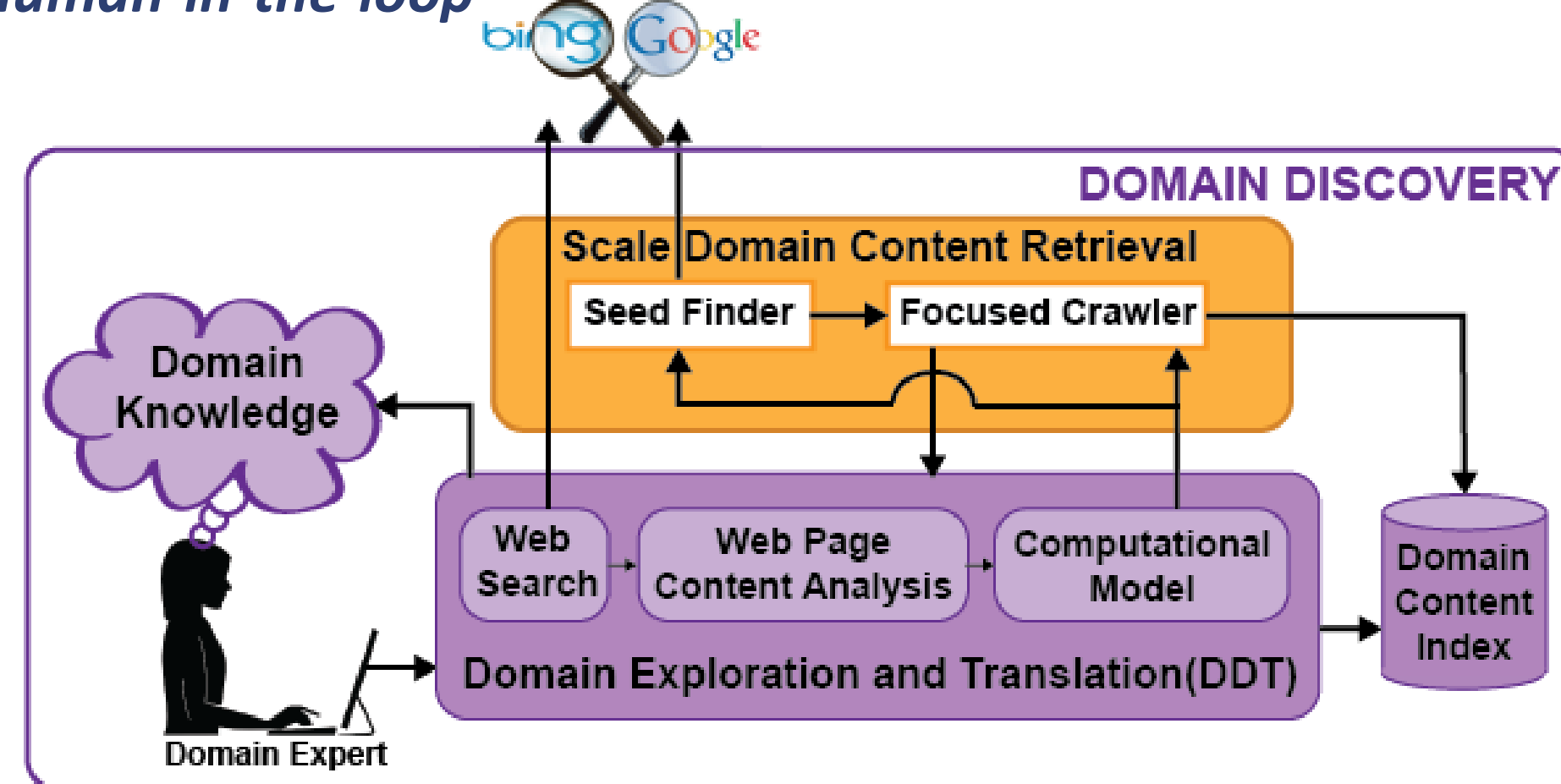
Memex Project: Fighting Against Human Trafficking

Yamuna Krishnamurthy, Kien Pham, Aecio Santos, Jorge Piazentin Ono, Sonia Castelo Quispe, Juliana Freire



DOMAIN DISCOVERY (DD) PROCESS

Iterative process to identify, retrieve and learn information and sources from the Web relevant for a specific information need with a human-in-the-loop



DOMAIN DISCOVERY DESIDERATA

- Support exploratory data analysis (EDA) of web pages
 - Multidimensional scaling visualization of pages
 - Maintain search context and capture analyst's feedback
 - Summarize search results
 - Multi-criteria filtering
- Translate the analyst's interactions with the Web pages into a computational model of the domain
 - Visualization for understanding the domain model
 - Further discovery of domain on the Web with model through:
 - Focused crawling (ACHE)
 - Automated searches (ACHE Seed Finder)

DOMAIN DISCOVERY TOOL

https://github.com/VIDA-NYU/domain_discovery_tool

COUNTERING HUMAN TRAFFICKING

GOAL: Help subject matter experts (SME) in NGO's trying to study the patterns of human trafficking find answers to such questions as:

"What is the average age found in escort ads with the phone number 6187098512?"

"Given a physical description of an individual, how many escort ads in a given region match that physical description?"

that could identify potential under-age victims.

PROBLEM: To answer these questions a SME has to gather content that contains such information, from hundreds of pages retrieved from the Web, to create a collection from which the required information such as age, phone numbers and locations can be extracted. This is a tedious process using existing Web tools.

SOLUTION: Use Memex Domain Discovery (DD) Tools (<https://opencatalog.darpa.mil/MEMEX.html>) to easily explore, understand, discover, gather and extract information from large number of pages from the Web that contain human trafficking related information.

DOCUMENT EXPLORER: Interpretable Multi-Dimensional Projection using Radviz

- Mass-spring force mechanism maps multi-dimensional data into a 2-dimensional visual space
- Identifies clusters and correlations between data
- Position of each point (page) in the 2-dimensional space is determined by the positions of the corresponding n-dimensional anchors (DA) which are words in the pages.
- Figure shows projection of pages for human trafficking domain

C-CHARTS: Visualizing and Understanding Computational Model Results

Id	Url	True label	Predicted label	Escort	Message	Others	Porn
19	https://www.carsfor...	Others	Others	0.000	0.000	1.000	0.000
75	http://www.autobyte...	Others	Others	0.000	0.000	1.000	0.000
72	https://www.carguru...	Others	Others	0.000	0.000	1.000	0.000
13	http://www.cars.com...	Others	Others	0.000	0.000	1.000	0.000
36	https://www.nadagu...	Others	Others	0.000	0.000	1.000	0.000

- a) C-Compare allows comparison of models
- b) C-Summary shows details of the selected models
- c) Visualization of per class model prediction that indicates correctly and incorrectly classified samples
- d) C-Explain provides an explanation for the selected samples through the words (features) contained in the samples using a word cloud.