MICHEL BIEZUNSKI, MATT NISHI-BROACH

TOPIC MAPS AT WORK

Semiotics Web / Information Architecture / OWASP Meetup Mashup

New York, March 30, 2016

WHO WE ARE

- Michel Biezunski, Infoloom Inc., once called "Joe Topic Map", has been the initiator and one editor of ISO/IEC 13250 Topic Maps. He has implemented Topic Maps for Encyclopedias, Government and Libraries.

- Matt Nishi-Broach is an artist, developer, and the creator of Axiologue, an ethical shopping initiative.

FINDING AIDS: PROS AND CONS

- **Tree-based**

   Taxonomies and Ontologies.

- **Graph-based**

   Linked Data, Google Knowledge Graph. RDF/Semantic Web.

- **Search algorithms**

   Trade secrets.

- **Crowd-sourced Tagging**

   Open, but messy over time.

BIG DATA / SMALL DATA

BIG	SMALL
DATA IS AVAILABLE BUT ITS CONTENT IS UNKNOWN.	PEOPLE KNOW WHAT THEY HAVE AND WANT TO SHOW IT.
DATA ANALYSIS OF HIGH VOLUME OF INFORMATION	FOCUS ON HIGH VALUE OF INFORMATION

TOPIC MAPS

- Combining open, automatable, technology-based information mining with user-controlled manual curation.

- Enable capturing of complex, ambiguous, multi-lingual information.

- Mess is not a bug, it's a feature !

WHERE TOPIC MAPS FIT

- Publishing / Media / Museums

- Industry / Technical Docs

- Library Reference

- Health Care

- Law

- Finance

- Government

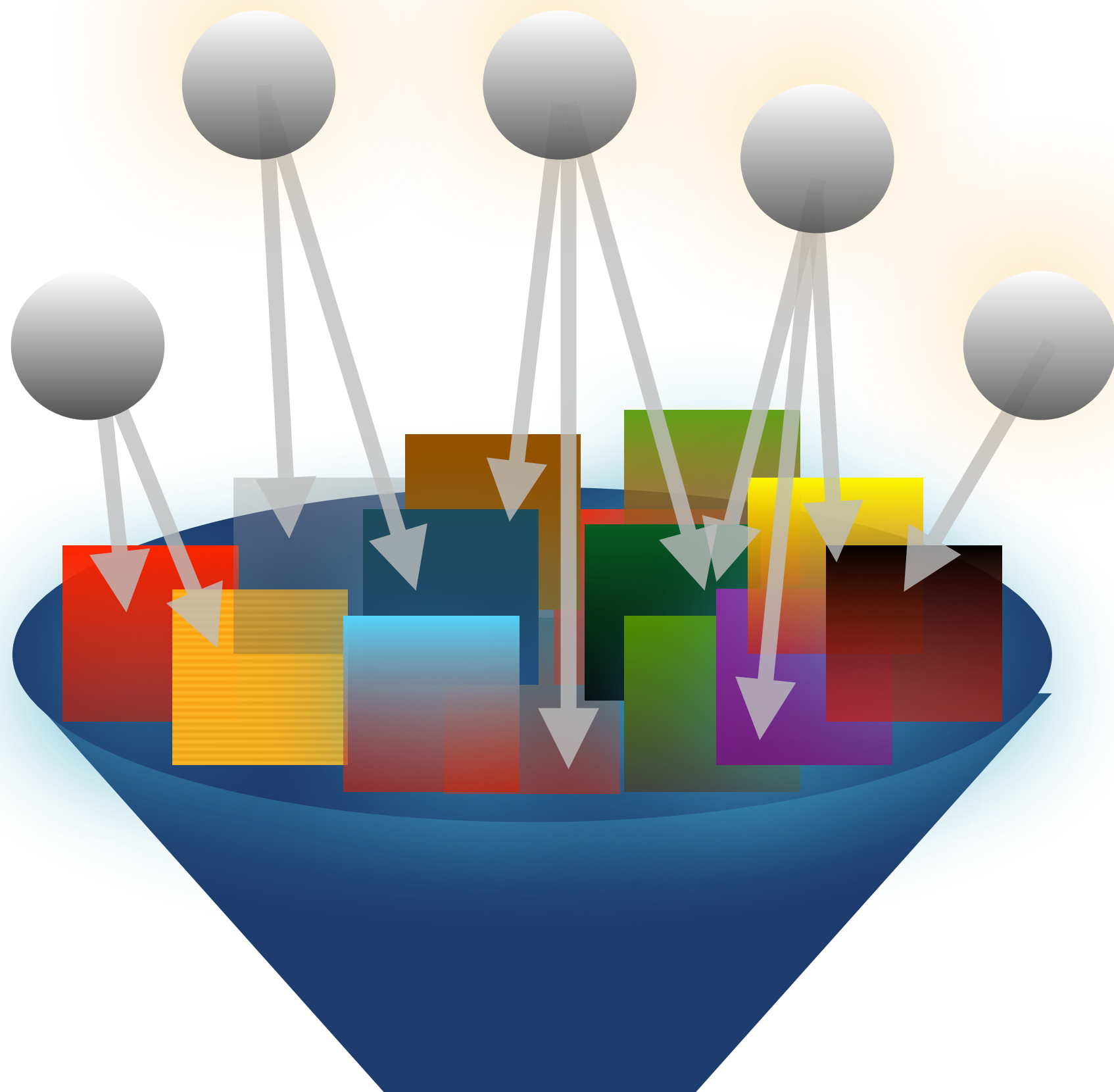- and … Curating results of Big Data Processing.

INTRODUCTION TO TOPIC MAPPING

- A topic is a computer representation of the knowledge available about any subject of conversation.
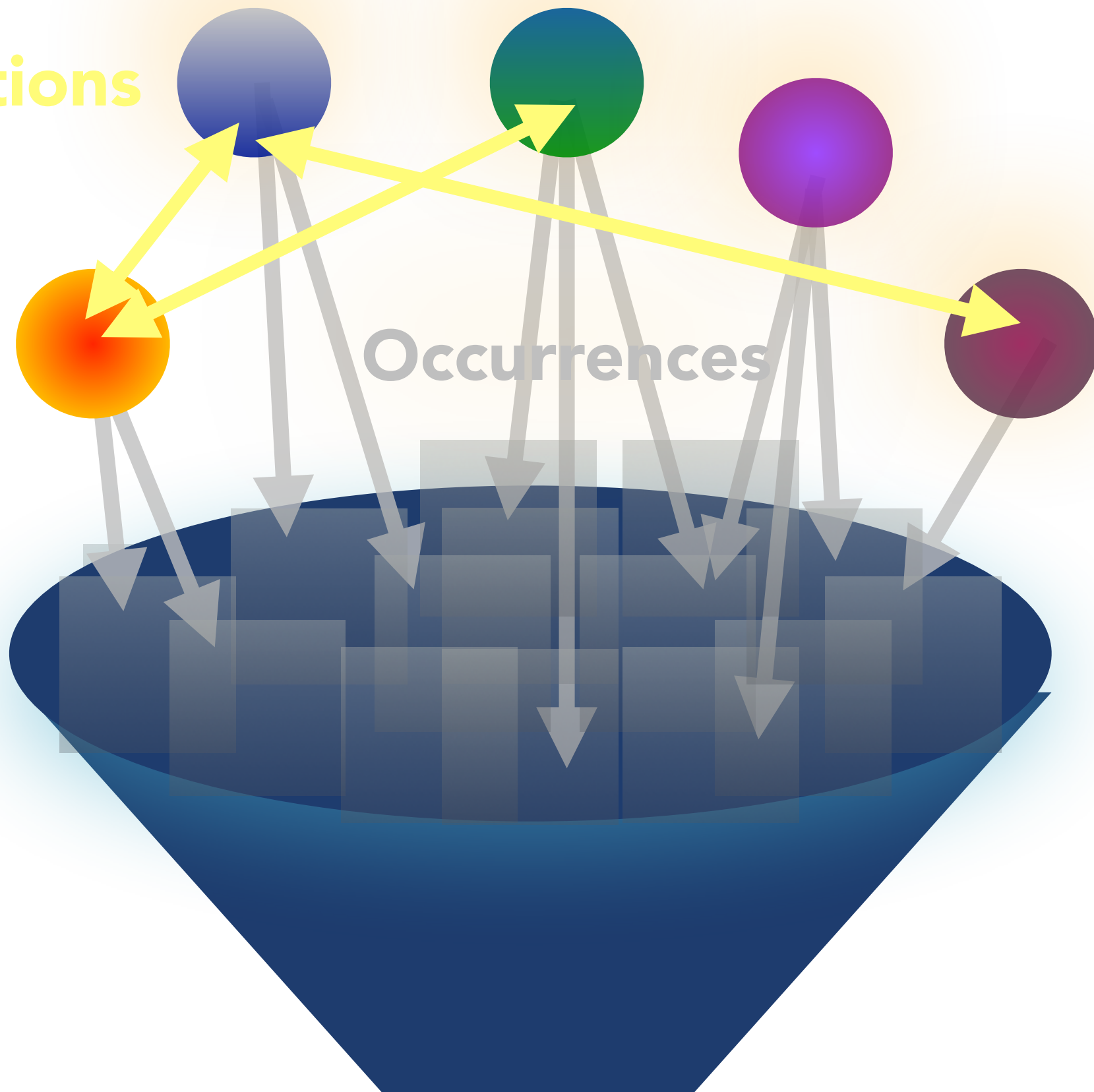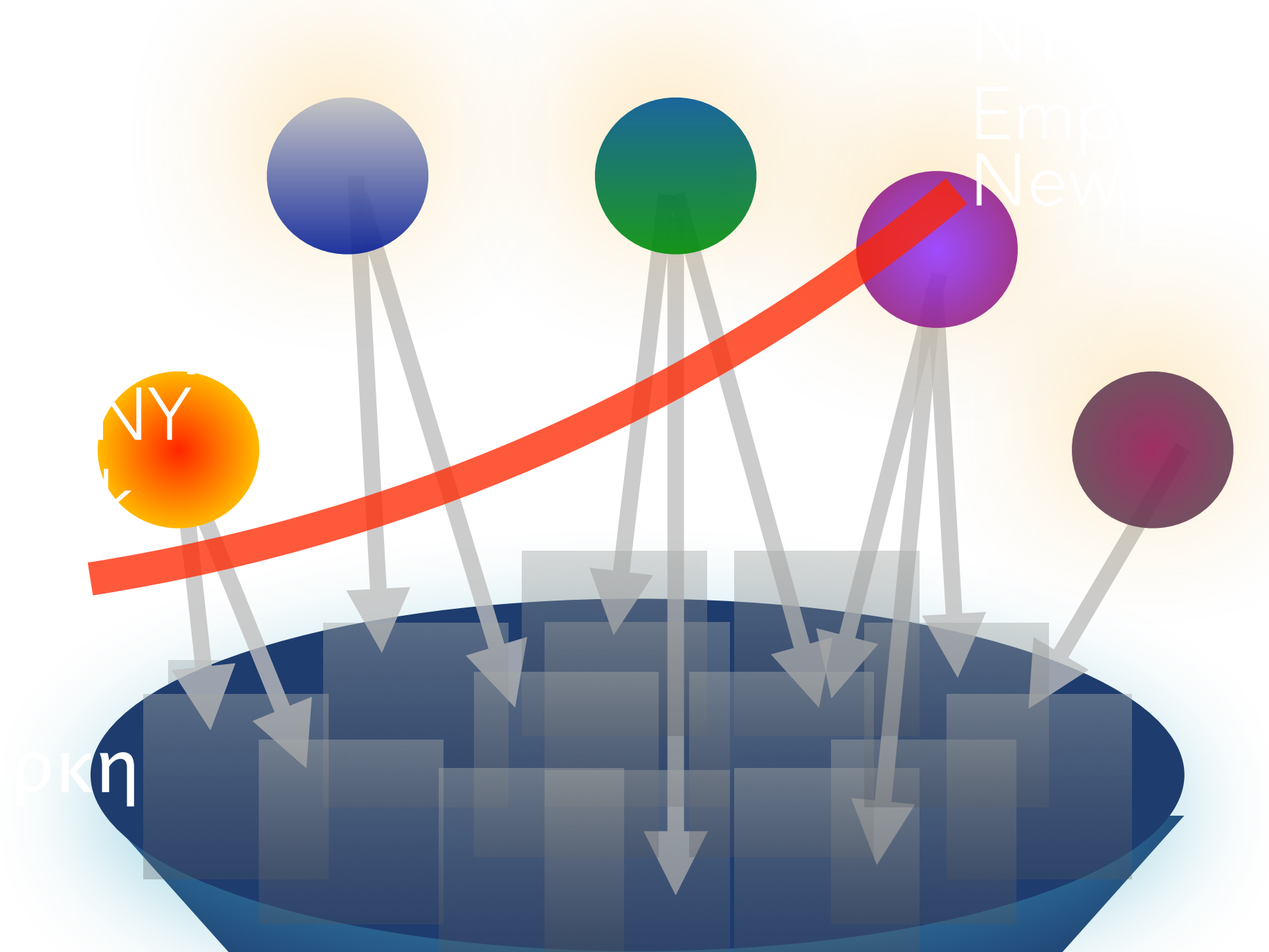
- A topic map is a graph of interconnected topics.

Relations

Topic

Names
Types

Occurrences

One name can be attached to multiple topics.

AMBIGUITY

New York (disambiguation)

From Wikipedia, the free encyclopedia

New York is a state in the United States of America.

New York, City, State, or other, may also refer to:

Places [edit]

England [edit]

- New York, Lincolnshire
- New York, North Yorkshire
- New York, Tyne and Wear

United States [edit]

In the state of New York

- New York City, a city in the State of New York and the most populous city in the United States
- New York metropolitan area, the region encompassing New York City and its suburbs
- New York County, covering the same area as the New York City borough of Manhattan
- Province of New York, a British colony preceding the state of New York

Elsewhere

- New York, Iowa, an unincorporated community
- New York, Kentucky, an unincorporated community
- New York, Texas, a community in Henderson County, Texas
- New York Mills, Minnesota, a city in Otter Tail County
- West New York, New Jersey, a town across the Hudson River from New York City, New York
- New York of the Pacific and New York Landing, former names of Pittsburg, California
- New York of the South, a nickname for Atlanta, Georgia
- New York, Missouri, a former community founded in 1835 or 1836 near Commerce, Missouri.

Black Sea [edit]

- The former Anglo-Saxon settlement of New York

Media and entertainment [edit]

DISAMBIGUATORS ARE OPTIONAL

NAME	SCOPE
NEW YORK	CITY
NEW YORK	STATE
NEW YORK NEW YORK	
NEW YORK TEXAS	

UNTIL THEY ARE NOT

NAME	SCOPE
NEW YORK NEW YORK	POSTAL ADDRESS IN MANHATTAN
NEW YORK NEW YORK	MOVIE
NEW YORK NEW YORK	THEME OF THE MOVIE NEW YORK NEW YORK
NEW YORK NEW YORK	SONG IN MUSICAL "ON THE TOWN"

CONSISTENT SCOPES DON'T ALWAYS WORK

NAME	SCOPE
NEW YORK CITY SUBWAY 14TH STREET STATION	6TH AVE LINE
NEW YORK CITY SUBWAY 14TH STREET STATION	7TH AVE LINE
NEW YORK CITY SUBWAY 14TH STREET STATION	8TH AVE LINE
NEW YORK CITY SUBWAY 14TH STREET STATION	BROADWAY LINE UNION SQUARE
NEW YORK CITY SUBWAY 14TH STREET STATION	14TH STR CROSSTOWN LINE

MULTILINGUAL TOPIC MAPS

- A topic can have multiple names in different languages.

- Filters can be set to only show the topics in a given language.

TAXONOMICAL TOPIC MAPS

- Topics can be distinguished by their types.

- Filtering is possible to get specialized lists by types:

  - e.g., Cities, People, Concepts, Historic events, Affordable Care Act, etc.

- Typed Topic Lists correspond to Specialized Indexes, Tables of Authorities, Product Catalogs, etc.

TOPIC MAPS ARE NOT NEW

- **Indexes** : List of topic names with occurrence indicators (in name alphabetic order)

- **Tables of Contents**: List of topics (headers), followed by their occurrences (in document order)

- **Cross-references**: Link between two (or more) occurrences of the same topic.

- **Glossaries, Dictionaries** : List of topic names and their occurrence content serving as definition.

- **Thesaurii**: Relations between topics

- **Library Catalogs, Web Taxonomies,** etc.

  Each of these finding aids is a pre-defined query in a topic map database.

OPERATIONAL TOPIC MAPS

- Knowledge is too complex to be handled just by automatic processes.

- Knowledge is too big to be handled entirely by hand (i.e. by brain).

- Design and find appropriate ways to customize automated knowledge acquisition, curation processes and publishing for the benefit of particular audiences.

WORKFLOW

- Aggregating

  - Import from structured (SQL, JSON, XML, MARC, etc.) or not structured data sets (Web, Plain text, Word, PDF, etc.)

- Curating

  - Merge/ Detach/ Hide/ Highlight/ Categorize/Relate/Multilingual

- Publishing

  - Push to Web, Mobile, Ebooks, PDF, etc. Live or Static applications.

DEMO

- OpenTopicLoom is being developed by Infoloom for its clients.

- Web-based Service to aggregate, curate and publish topic maps.

- It supersedes a technology which is being used for 15++ years and is still used to produce Tax Map for the IRS.