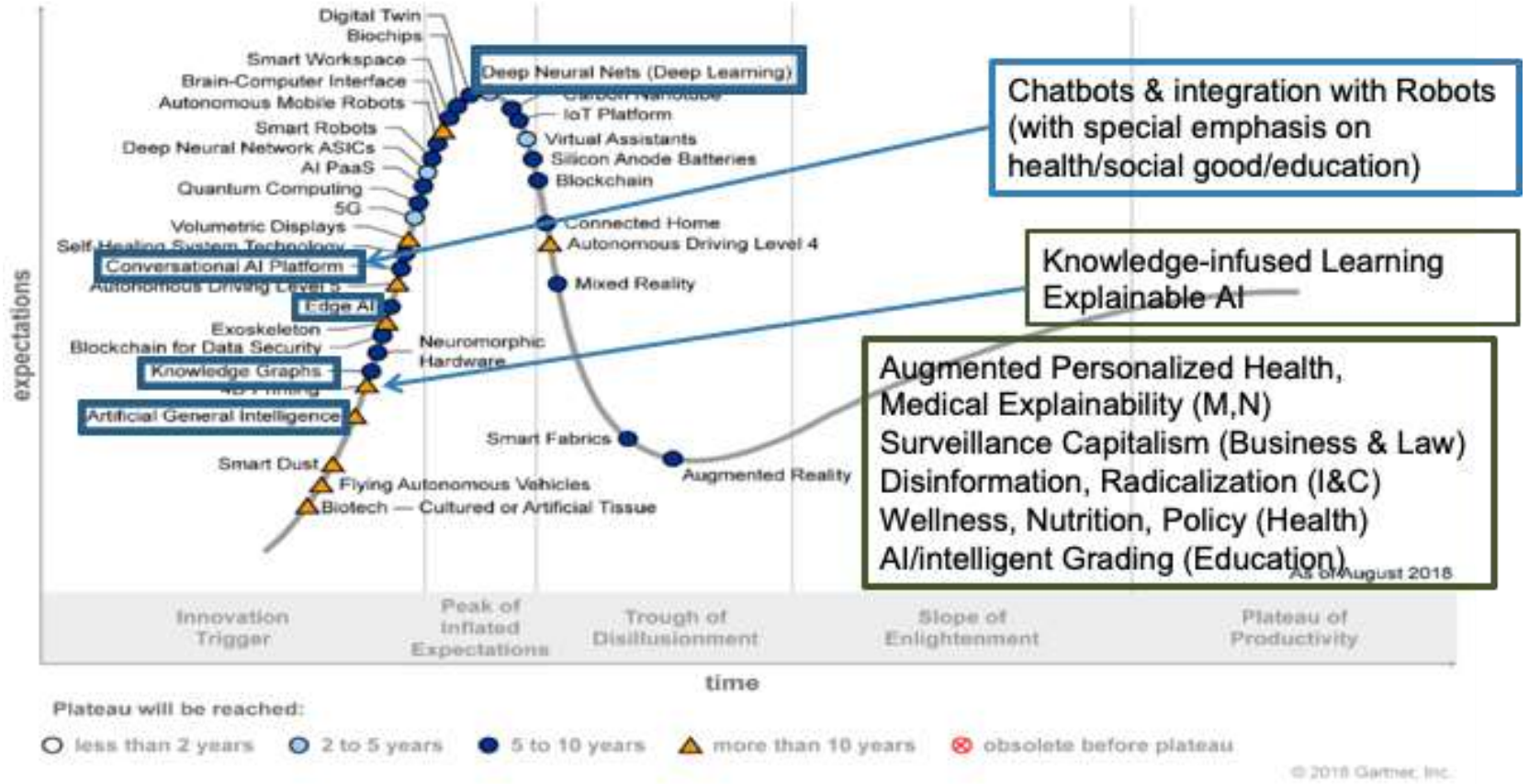
Knowledge Graphs and Their Central Role in Big Data Processing: Past, Present, and Future

7th ACM India Joint International
Conference on Data Science & Management of Data (CODS-COMAD)
Indian School of Business, Hyderabad Campus
5-7 January 2020

Prof. Amit Sheth
http://ai.sc.edu/

Artificial Intelligence Institute
University of South Carolina

AI Institute @ UofSC & AI Landscape



Interdisciplinary AI: Sample Collaborations

College of Arts and Sciences: on misinformation and radicalization (with the Institute of Mind and Brain, Comp Sc & Engg): a MURI proposal submitted (budget $6.25million) o recovery and resilience from natural disasters (with Hazards and Vulnerability Research Institute, ext partners)

College of Pharmacy (includes the South Carolina Colon Cancer Prevention Network): early onset of colorectal cancer on Digestive Inflammation Index

College of Education: develop personalized education plan and Analogy based learning in undergraduate courses (eg, biochemistry): using chatbots (virtual agents) for the health of school children through nutrition monitoring

Mech Engineering - using chatbot/AI in 2+2 educational in connected manufacturing

Collaborations...continued

School of Medicine and College of Nursing: personalized digital health with virtual assistants (chatbots) for (a) pediatric patients with neutropenia, (b) asthma in children, (c) obesity and hypertension in adults (d) mental health in adults

Arnold School of Public Health: o brain imaging (with Aging Brain Cohort and Aphasia Laboratory), o brain aneurysm/ strokes o identification and management of mental health including depression and suicide ideation

School of Journalism and Mass Communications: social media insights and harassment of journalists (with Social Media Insights Lab)

Knowledge - critical role in intelligent computing

Knowledge - identifying, representing, capturing/creating, representing: ontology, knowledge graph

Associating meaning to data, making data interpretable: e.g., semantic annotations

Supporting planning, reasoning, analysis, insight, decision making: semantic applications (search, browsing, ….), improving/enhancing NLP and ML, to activities that support humans

With massive growth in Big Data, this talk is about the central role of Knowledge in computing - in the past, present and future

The role of knowledge in computing has long been recognized - at least since Vannever Bush's 1945 seminal piece: ***As We May Think****.*

Knowledge Graphs (KG)

is a structured knowledge in a graphical representation.

Knowledge Networks (KN)

integrate and combine knowledge (usually captured as KGs) from various domains.

Enhanced (semantic) applications such as search, browsing, personalization, recommendation, advertisement, and summarization.

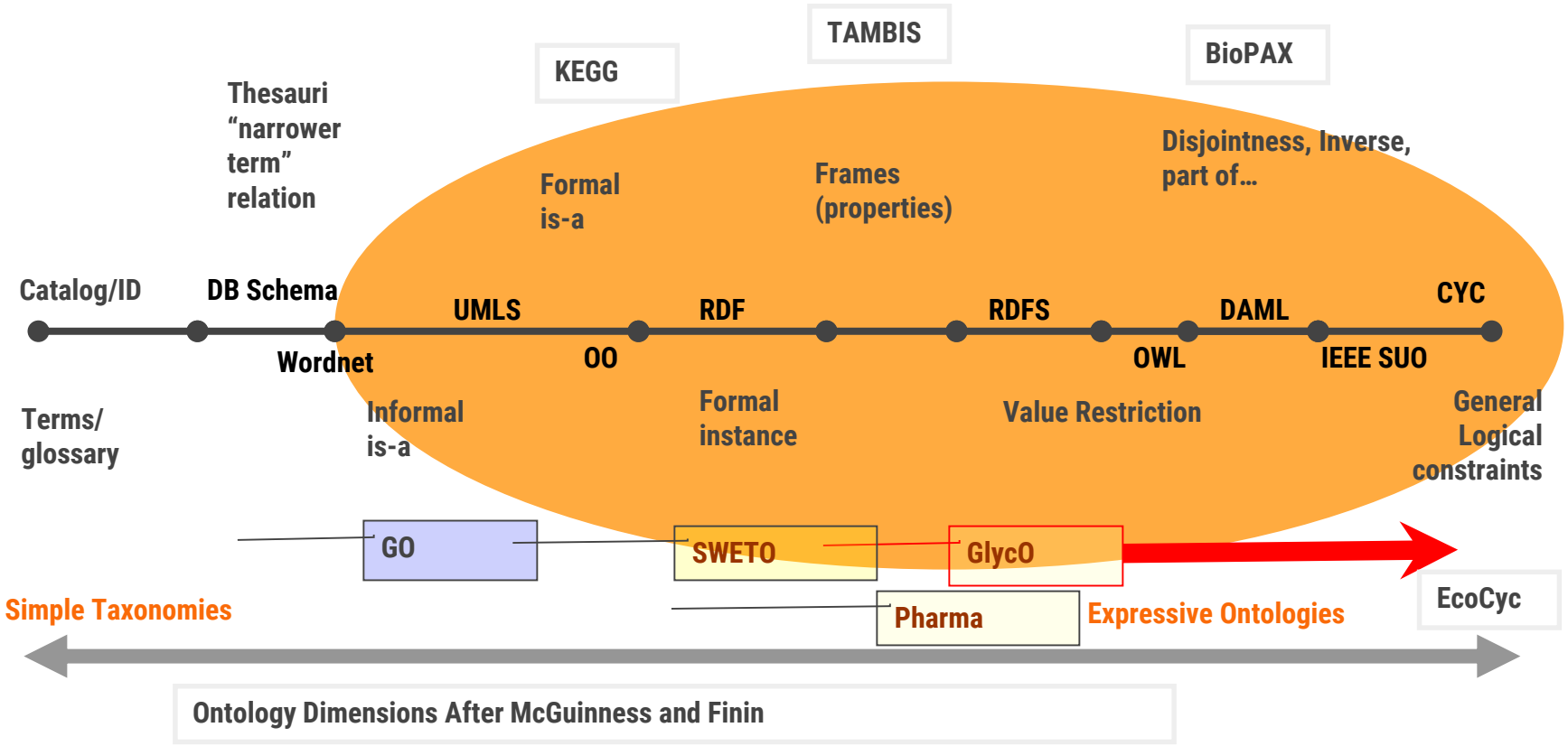
Improve integration of data, including data of diverse modalities and from diverse sources.

Empower/enhance ML and NLP techniques. Use as a knowledge transfer mechanism across domains, between humans and machines

Improve automation and support intelligent human-like behavior and activities that may involve conversations or question-answering and robots.

Expressiveness Range:

Knowledge Representation and Ontologies



Forms of KR pursued

- Simple Facts/Assertions
- Taxonomic
- Schema (description)/instances (description base)
- Logical/multiclass - DL, FOL, higher order
- Graphs - Relationships (non-directional/directional, unlabeled/labeled, and constraints)
- Probabilistic Graphs

What about statistical (i.e., implicit knowledge in statistical AI)?

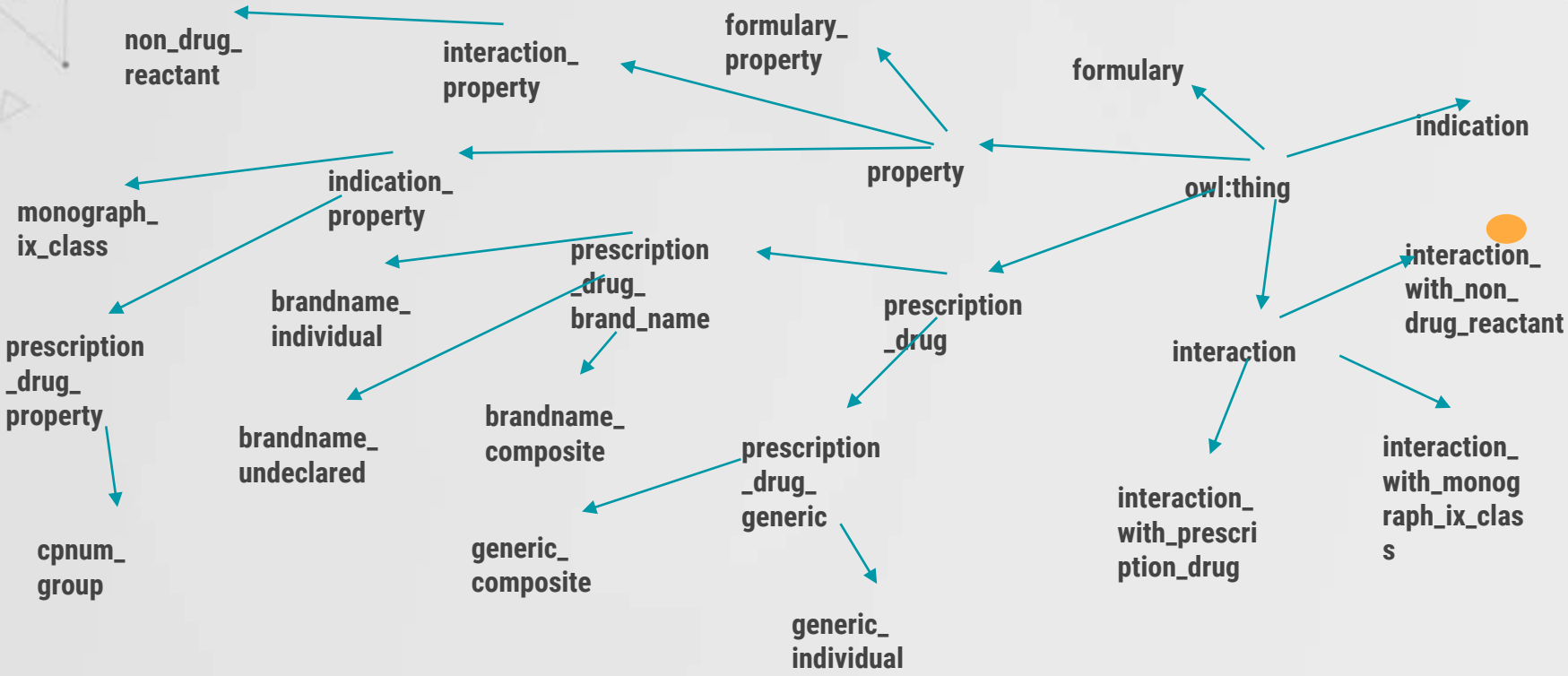
Semantic Web – Intelligent Content

(supported by Taalee Semantic Engine)

Figure: *Semantic Web - Semagix*

Drug Ontology Hierarchy

(showing is-a relationships)



Multimodal KG

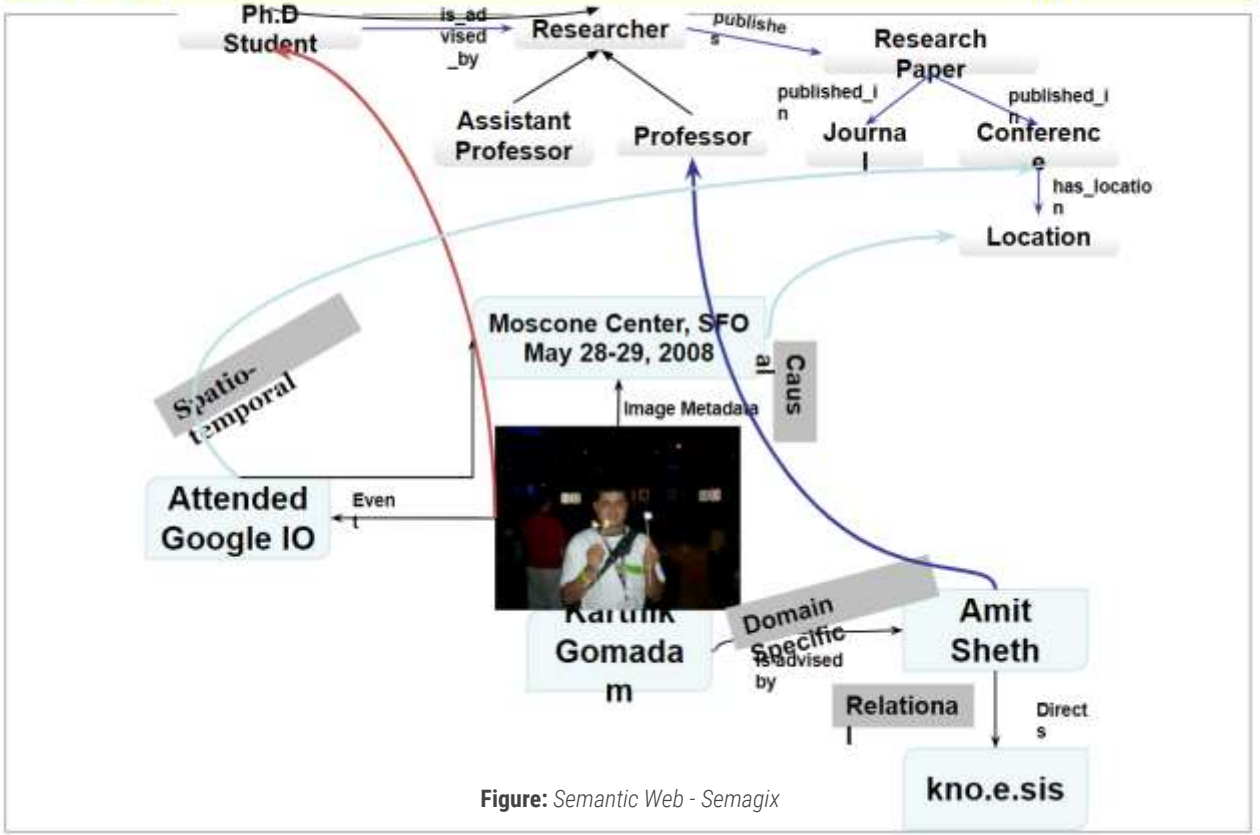


Figure: *Semantic Web - Semagix*

SEMAGIX
POWER • THROUGH • RELEVANCE

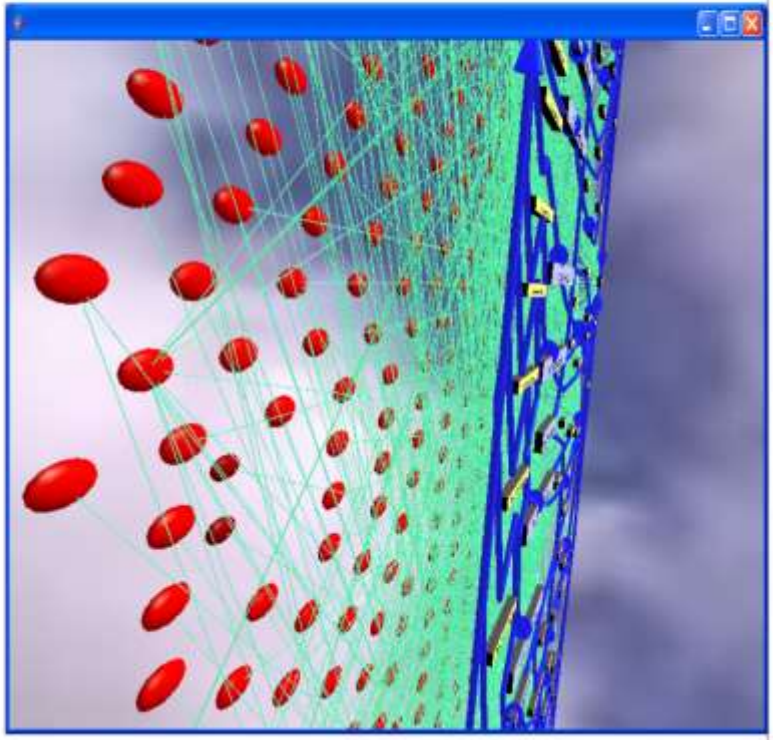
Space Partitioning

Foreground

- visualization of entities and their properties in the foreground.

Background

- documents are visualized in the background.



52

Figure: *Space Partitioning*

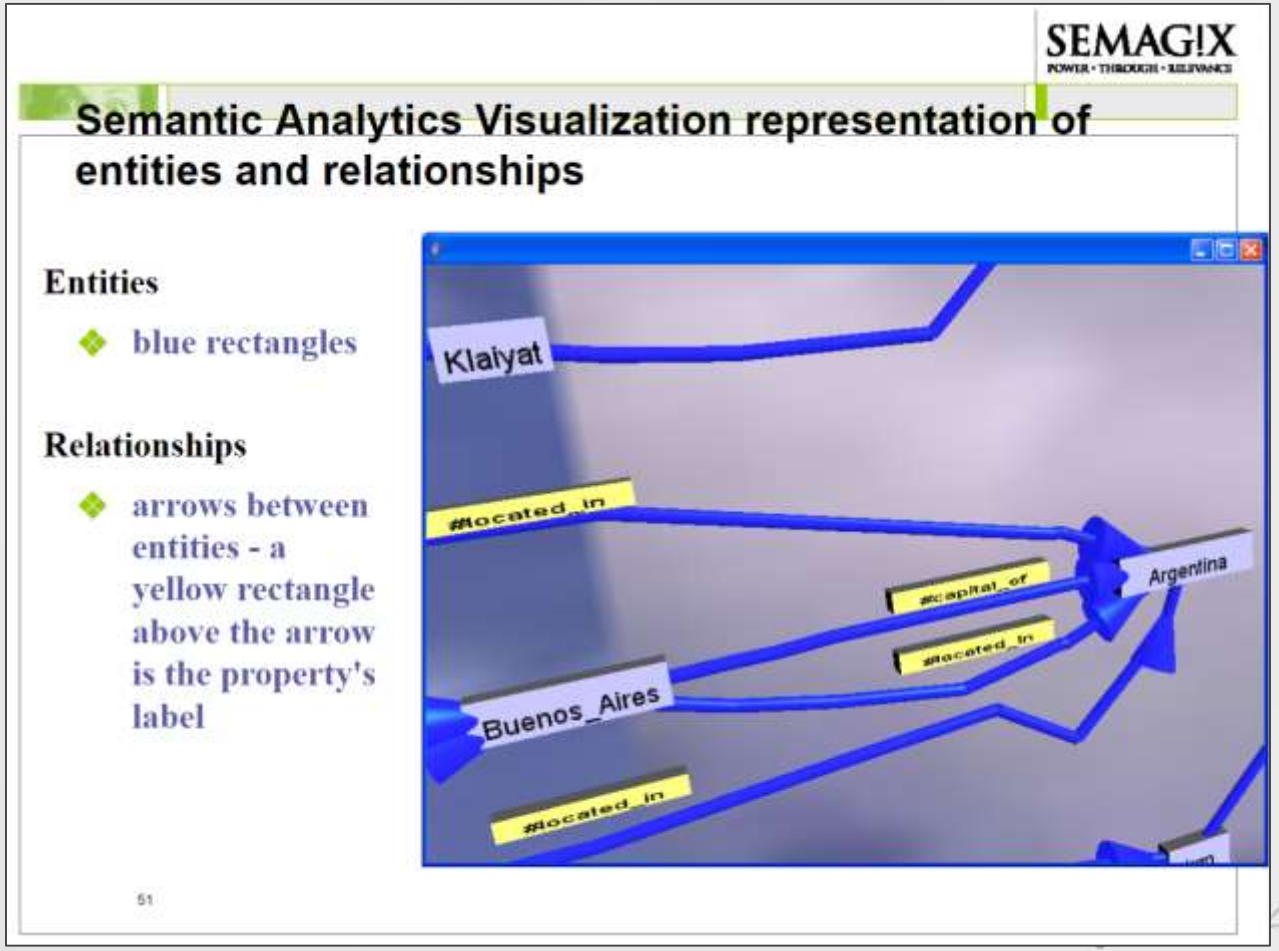


Figure: *Semantic Analytics & Visualization*

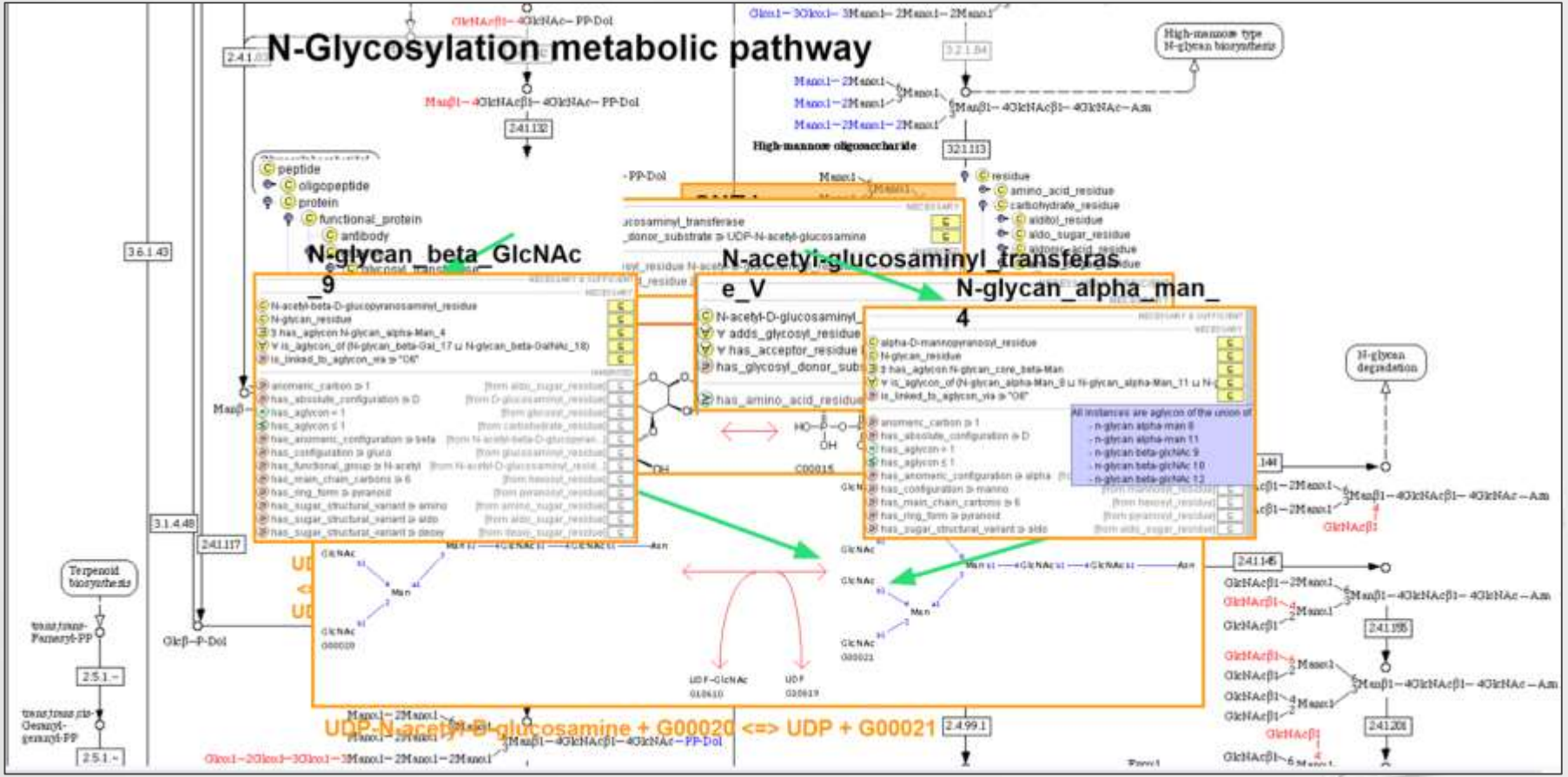
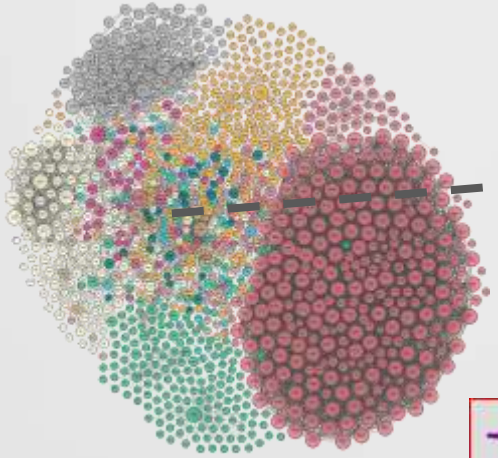


Figure: *N-Glycosylation Metabolic Pathway*

Knowledge Graphs have become prominent recently



Linked Open Data > 9960 datasets,
> 149 B triples

DBpedia

38.3 M entities and 8.8 B facts

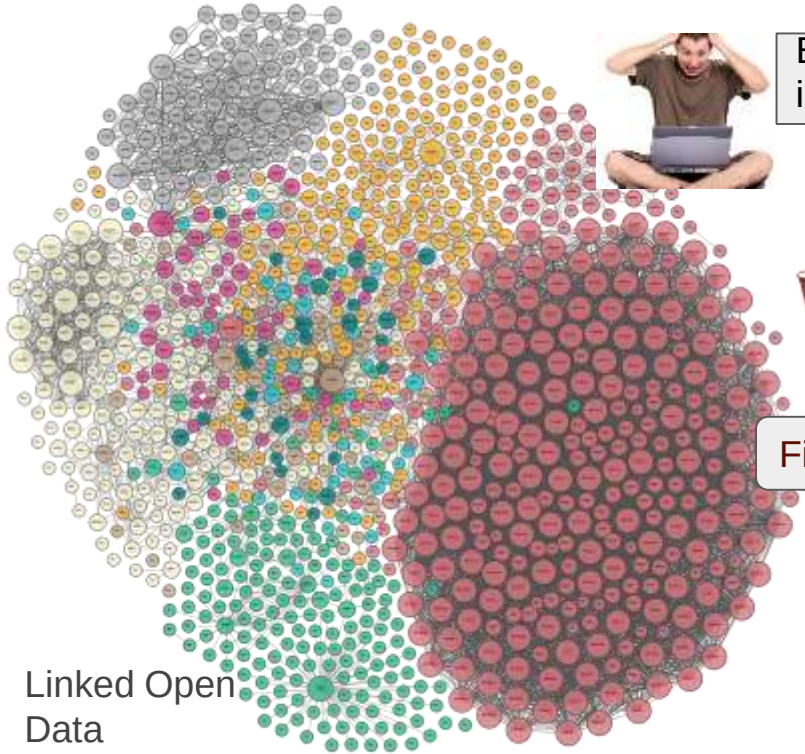
Google Knowledge Graph 570 M entities and 18 B facts

YAHOO!
Google = schema.org
bing

Schema.org annotations

LinkedIn knowledge graph

Domain-specific knowledge extraction from LOD



Linked Open Data

Book related information?

Filter relevant datasets

Books, movie, games

DBTropes

Books, Countries, Drugs

DBpedia

Books

Project Gutenberg

Extract relevant portion of a data set

Book specific DBpedia

Book specific DBTropes

http://knoesis.org/node/2272

http://knoesis.org/node/2793

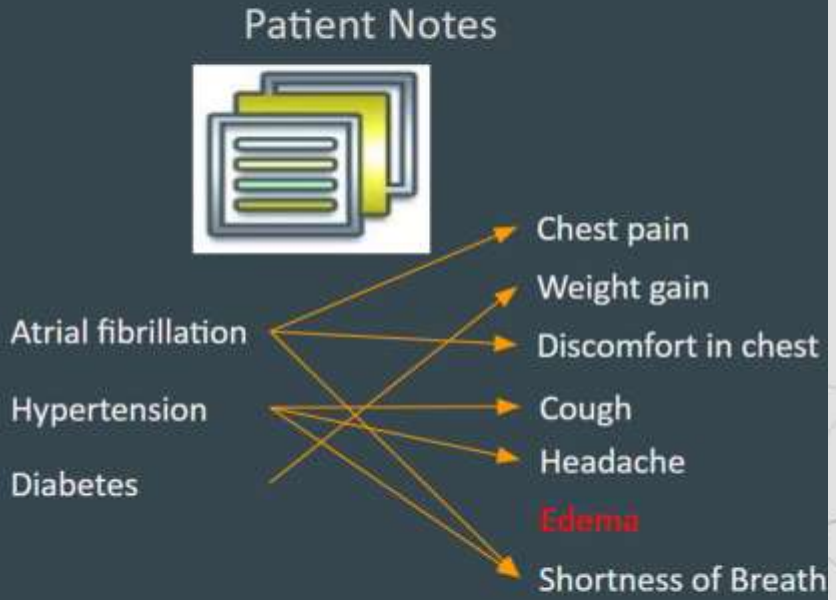
It is becoming easier to find or create relevant knowledge for a given application

Existence of large knowledge bases
Ability to search/find a relevant knowledge bases [WI'13]
Ability extract a relevant subset [IEEE Big Data'16]
Ability to enrich - by deriving new concepts and new facts [BIBM'12]

Knowledge graphs are already playing influential roles in many applications involving big data, starting with search [15 years of search & knowledge graphs].

Ability to enrich knowledge graphs

Initial knowledge graph on disorder and symptoms



Initial knowledge base does not know about edema. Can Edema be a symptom of any of the disorders mentioned according to the patient notes?

http://knoesis.org/node/2642

Int'l Journal on Semantic Web & Information Systems, 1(1), 1-18, Jan-March 2005

Semantics for the Semantic Web: The Implicit, the Formal and the Powerful

Amit Sheth, Cartic Ramakrishnan and Christopher Thomas
Large Scale Distributed Information Systems lab
University of Georgia, Athens, GA USA

Abstract:

Enabling applications that exploit heterogeneous data in the Semantic Web will require us to harness a broad variety of semantics. Considering the role of semantics in a number of research areas in computer science, we organize semantics in three forms: implicit, formal and powerful, and explore their roles in enabling some of the key capabilities related to the Semantic Web. The central message of this paper is that building the Semantic Web purely on description logics will artificially limit it potential, and that we will need to both exploit well known techniques that support implicit semantics, and develop more powerful semantic techniques.

Keywords: Semantic Web, Semantic Technology, Formal Semantics, Informal Semantics, Implicit Semantics, Analytical Processing, Document Management and Retrieval, Knowledge Discovery, Soft Computing, Metadata, Semantic Search, Relationship Discovery, Semantic Analytics, Semantic Matching, Semantic Integration

Sheth, A., Ramakrishnan, C., & Thomas, C. (2005). Semantics for the semantic web: The implicit, the formal and the powerful. International Journal on Semantic Web and Information Systems (IJSWIS), 1(1), 1-18.

What propels popularity of KG and its development?

KG enabled Web and Enterprise Applications: *Microsoft, Siemens, LinkedIn, Airbnb, eBay, and Apple, as well as smaller companies (e.g. ezDI, Fraanz, Metaphactory/Metaphacts GmbH, Semantic Web Company GmbH, Mondeca, Stardog, Diffbot, Siren)*

Enhanced Learning: *"Data alone is not enough (Pedro Domingos)"* and the use of domain knowledge improves the results or effectiveness of state of the art ML and NLP techniques.

Linked Open Data (LOD)

Schema.org (schema.org)

Data Commons Knowledge Graph (DCKG) (datacommons.org)

Wikidata (wikidata.org)

Characteristics of Enterprise KG

	Data model	Size of the graph	Development stage
Microsoft	The types of entities, relations, and attributes in the graph are defined in an ontology.	~2 billion primary entities, ~55 billion facts	Actively used in products
Google	Strongly typed entities, relations with domain and range inference	1 billion entities, 70 billion assertions	Actively used in products
Facebook	All of the attributes and relations are structured and strongly typed, and optionally indexed to enable efficient retrieval, search, and traversal.	~50 million primary entities, ~500 million assertions	Actively used in products
eBay	Entities and relation, well-structured and strongly typed	Expect around 100 million products, >1 billion triples	Early stages of development and deployment
IBM	Entities and relations with evidence information associated with them.	Various sizes. Proven on scales documents >100 million, relationships >5 billion, entities >100 million	Actively used in products and by clients

Industry-Scale Knowledge Graphs: Lessons and Challenges (**Communications of the ACM, August 2019)**

PAST

Knowledge Graph for Semantic Applications

Early Use of Knowledge Graphs,

before the start of this century, related to building a knowledge graph **manually** or **semi-automatically** and applying them for *semantic applications*, such as search, browsing,
personalization, and advertisement.

Examples: Taalee, Semagix Semantic Search (2000), InfoBox

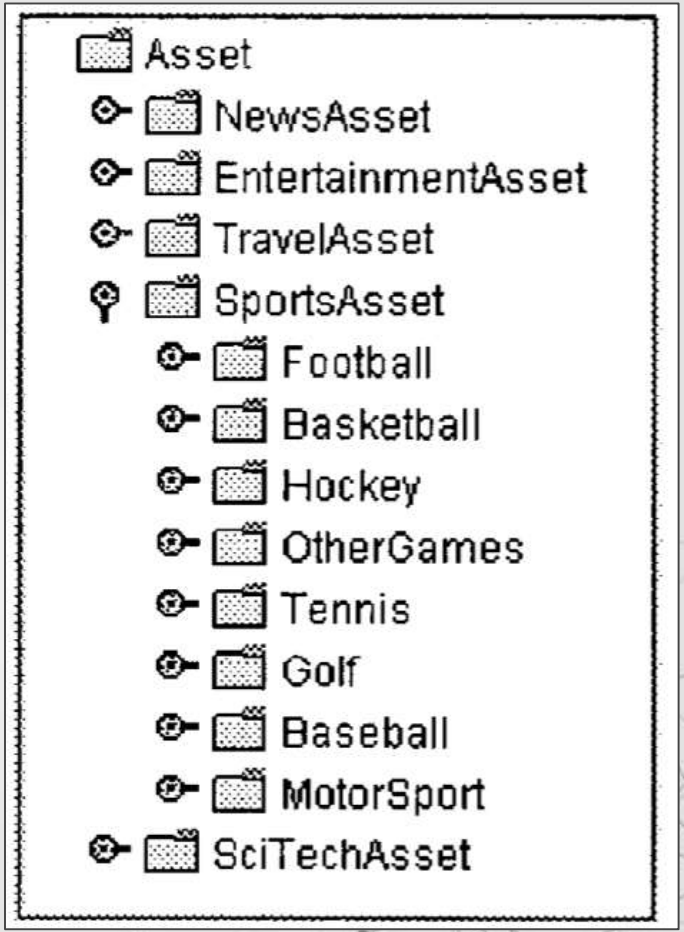
http://bit.ly/15yrSemS

Creation & Use of Knowledge ~2000

(12) **United States Patent**
Sheth et al.

(54) **SYSTEM AND METHOD FOR CREATING A SEMANTIC WEB AND ITS APPLICATIONS IN BROWSING, SEARCHING, PROFILING, PERSONALIZATION AND ADVERTISING**

(75) Inventors: **Amit Sheth**; **David Avant**, both of Bogart; **Clemens Bertram**, Athens, all of GA (US)



SEMAGIX

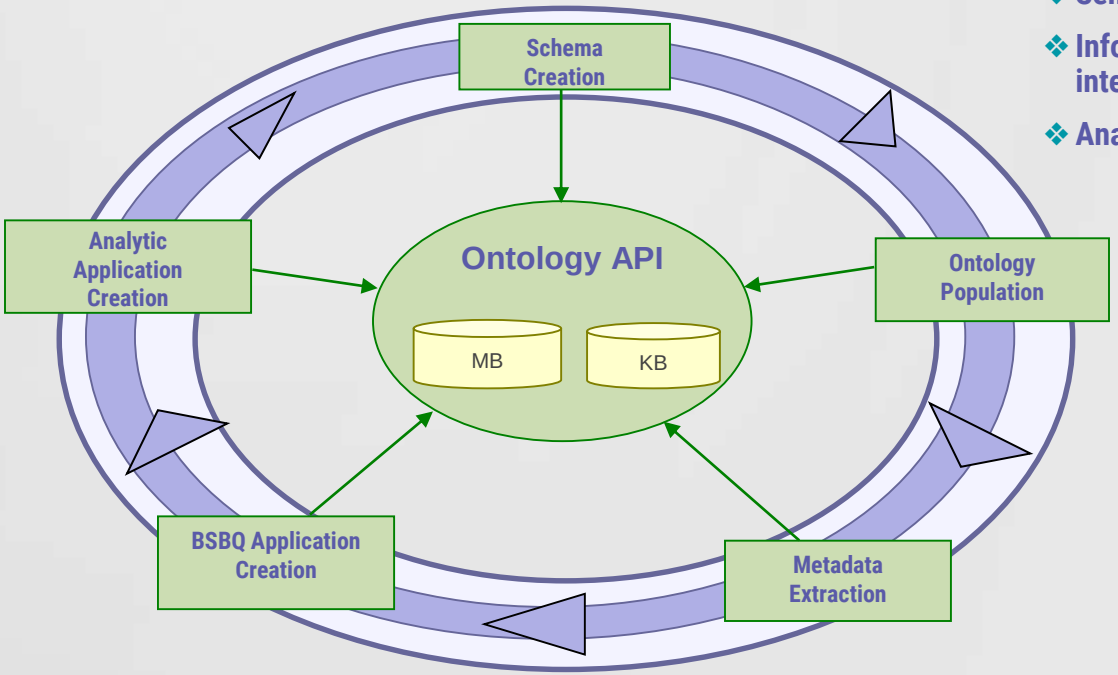
Power Through Relevance

Additional reading: **15 years** of Semantic Search and Ontology-enabled Semantic Applications (https://www.linkedin.com/pulse/15-years-semantic-search-ontology-enabled-amit-sheth/)

Semantic (Web) Technology
State of the Art

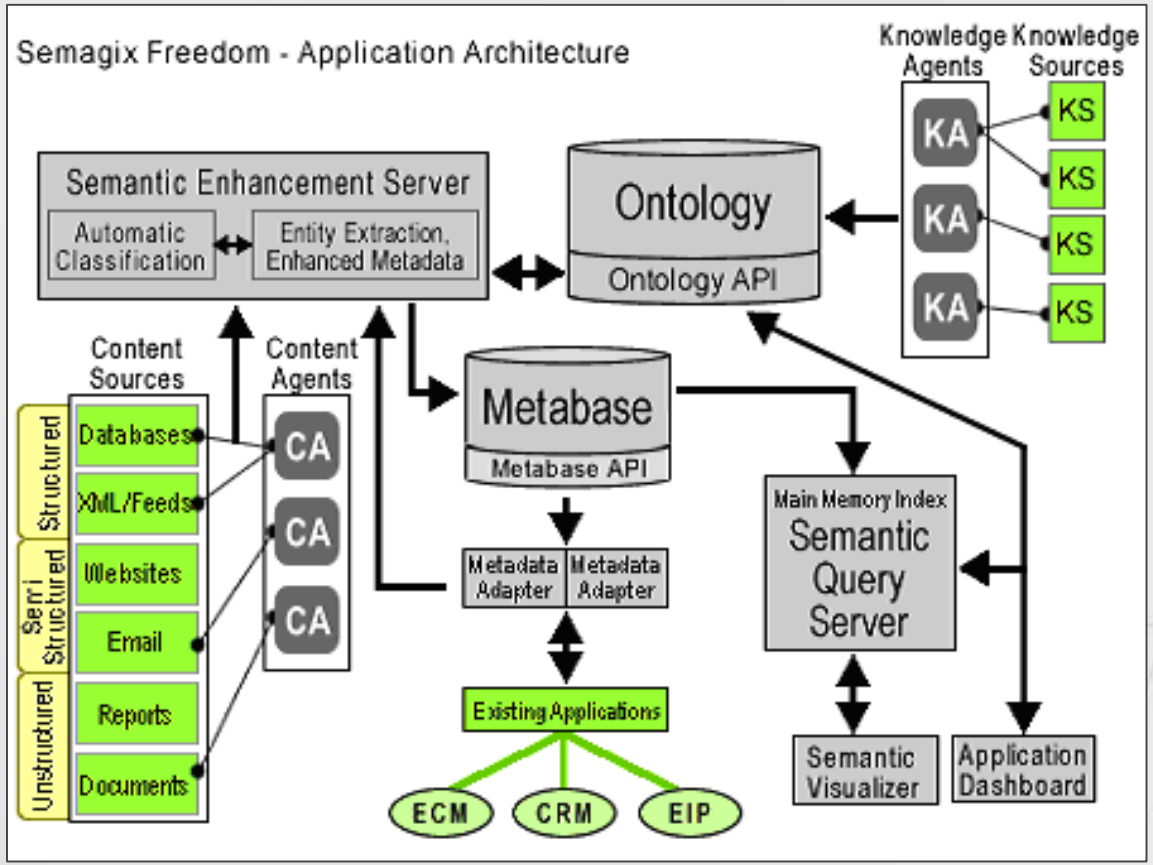
Building a scalable and high performance system with support for:
Ontology creation and maintenance
Ontology-driven Semantic Metadata Extraction/Annotation
Utilizing semantic metadata and ontology

- **Semantic search/querying/browsing**
- **Information and application integration - normalization**
- **Analysis/Mining/Discovery – relationships**



Semagix Freedom Architecture

for building ontology-driven information system



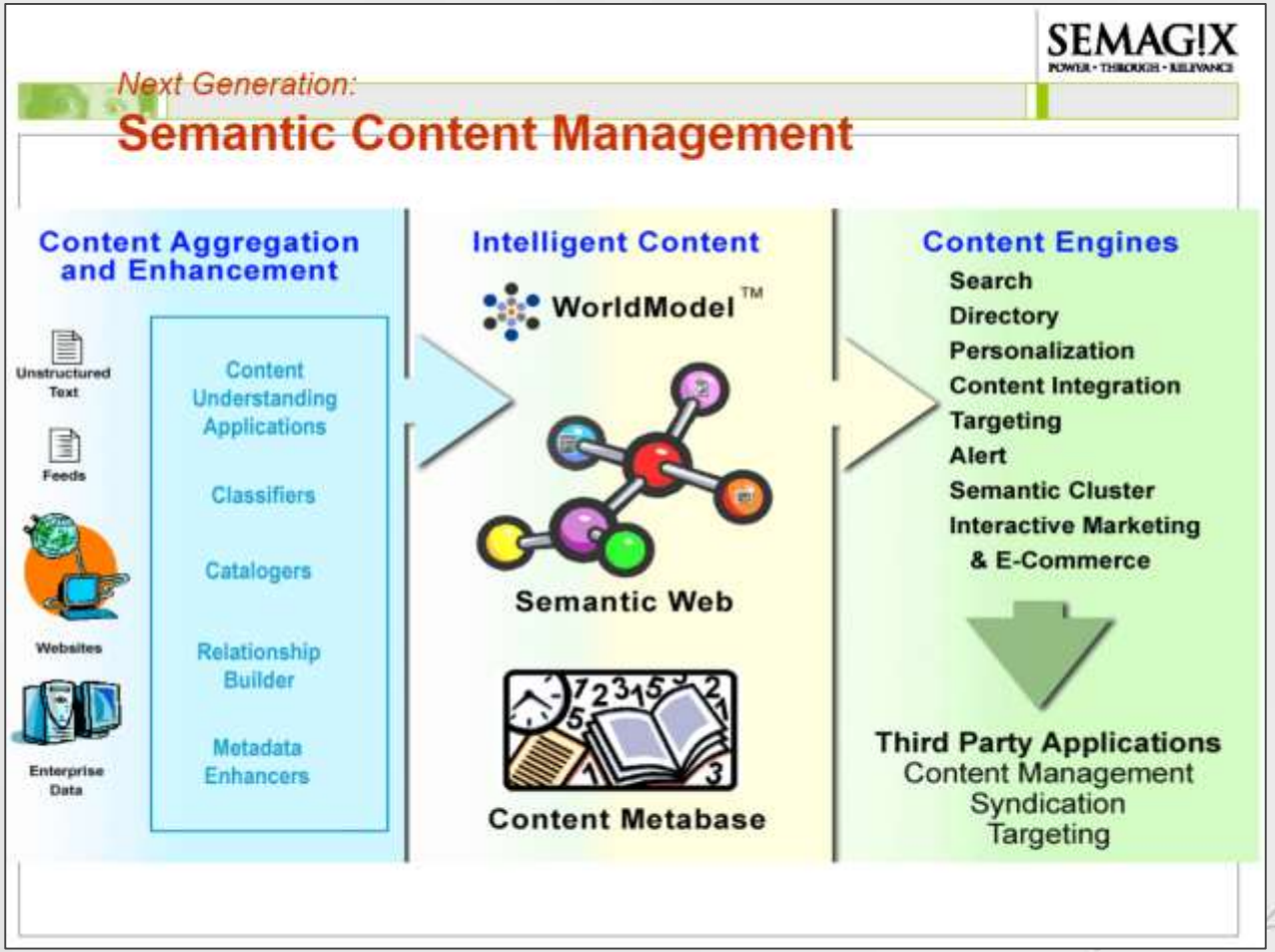
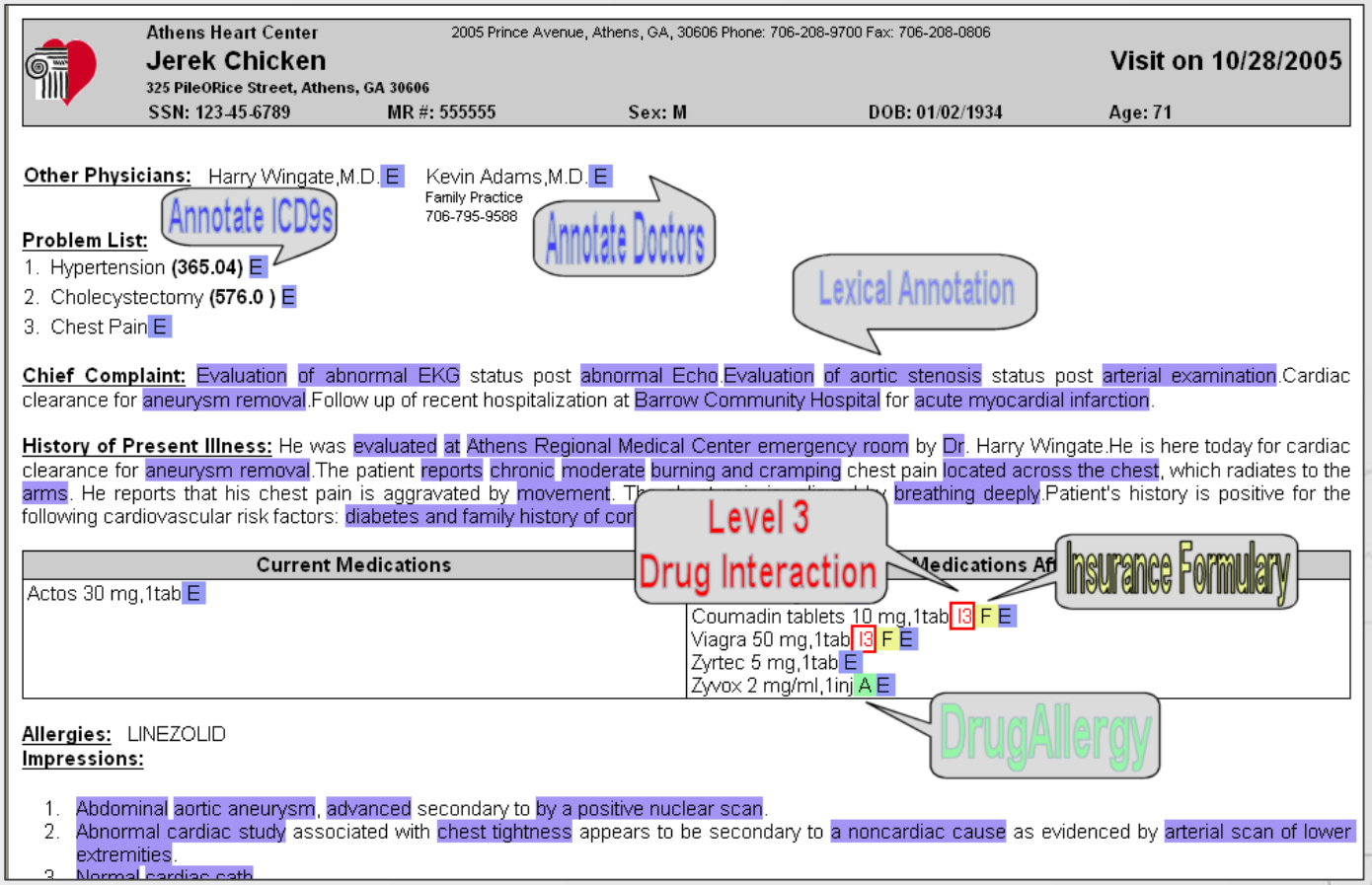


Figure: *Semantic Content Management - Semagix*

Semantic Annotation/ Metadata Extraction + Enhancement

Active Semantic Doc with 3 Ontologies



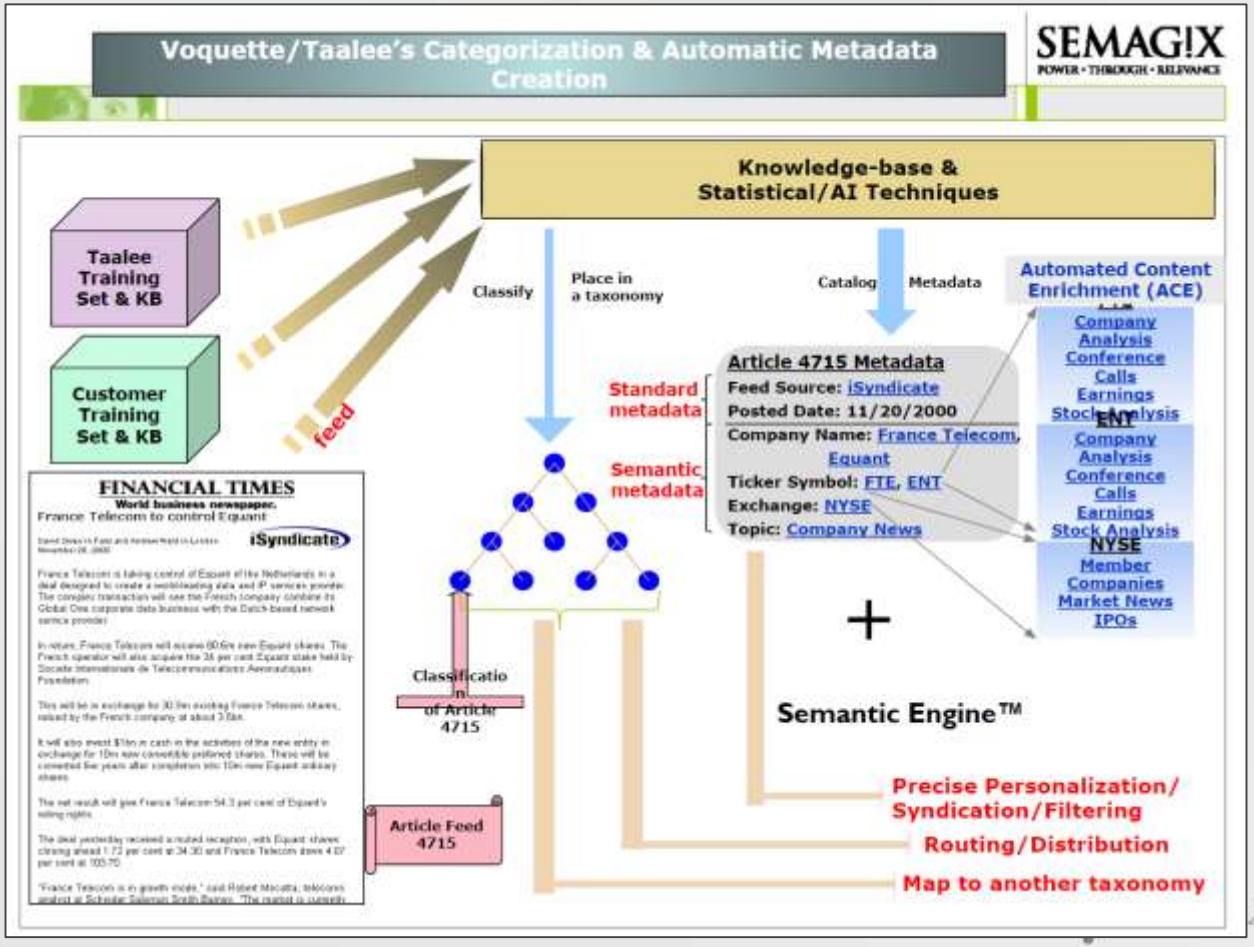


Figure: *Semantic Enhancement Engine - Taalee*

Figure: *Semantic Search - Semagix*

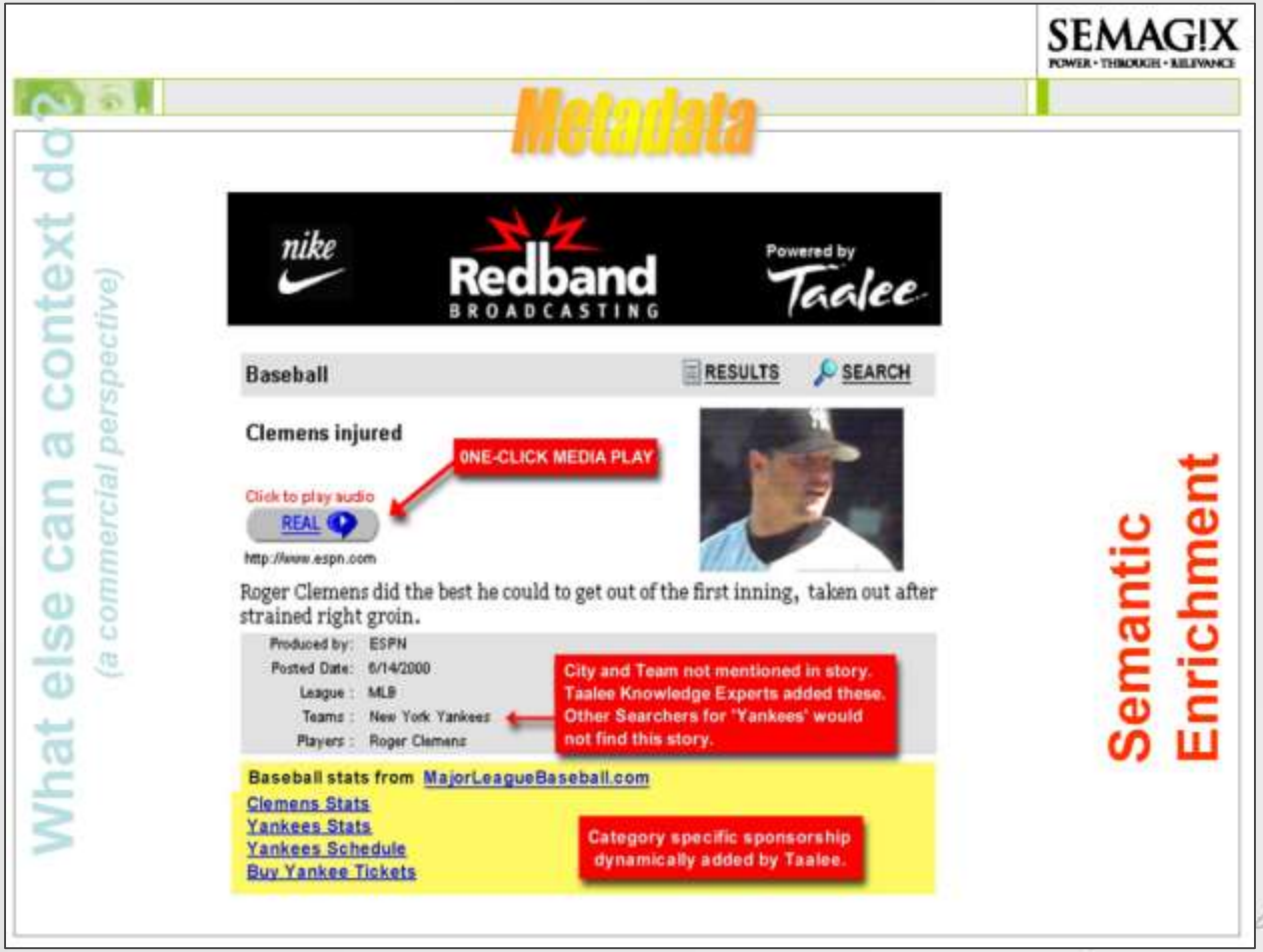


Figure: *Semantic Metadata - Semagix*

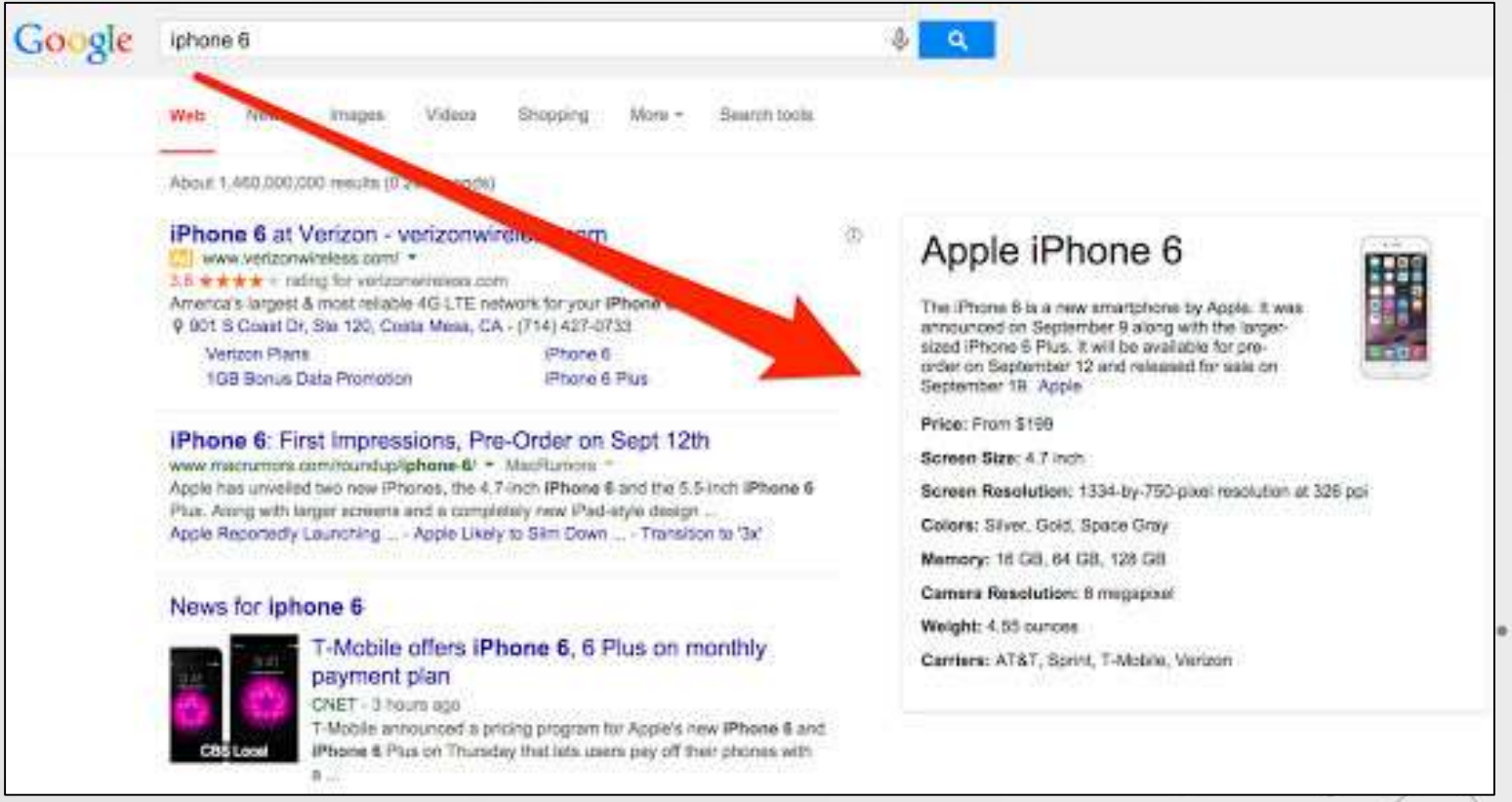


Figure: *Google InfoBox*

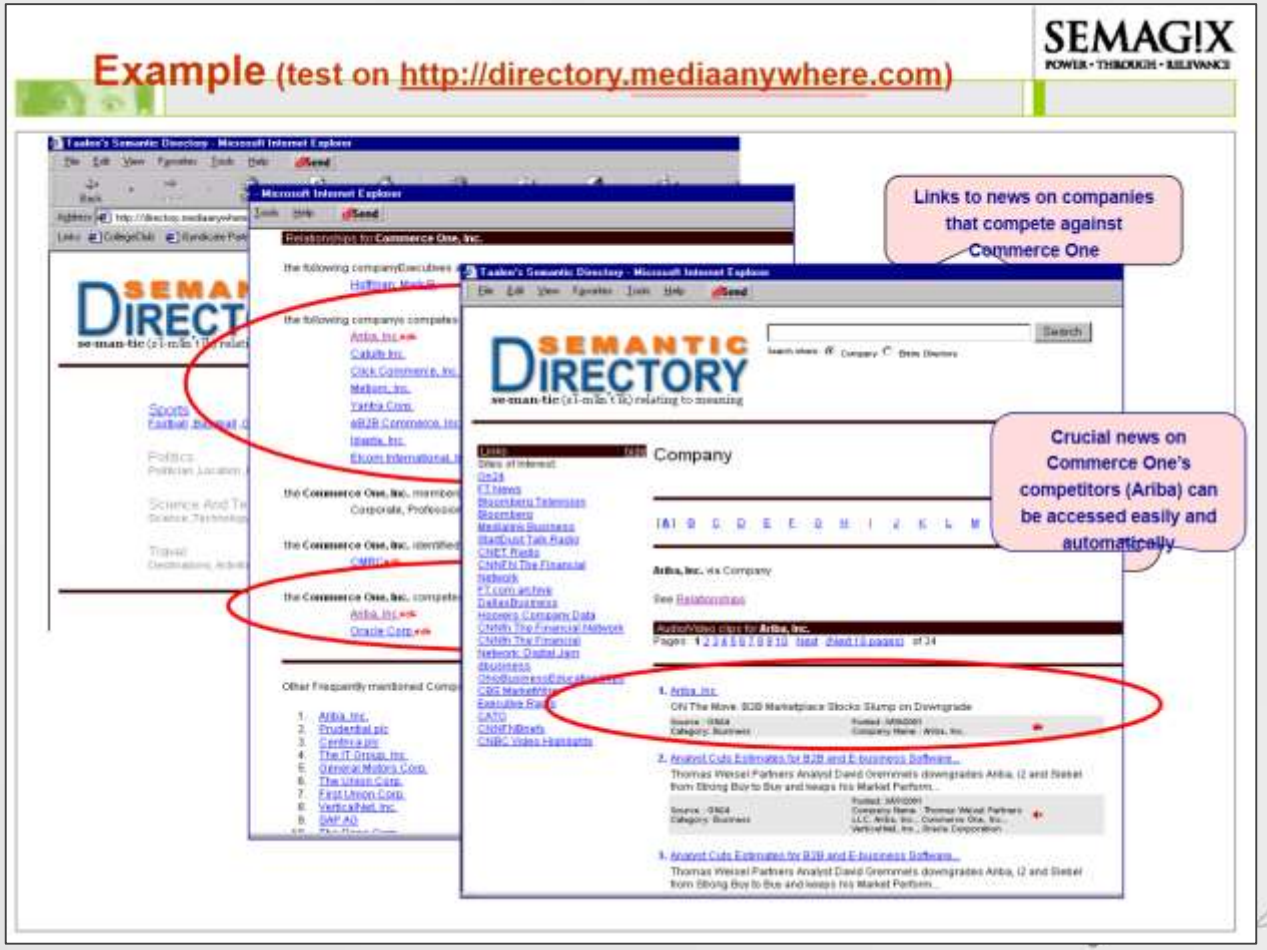


Figure: *Semantic Directory - Semagix*

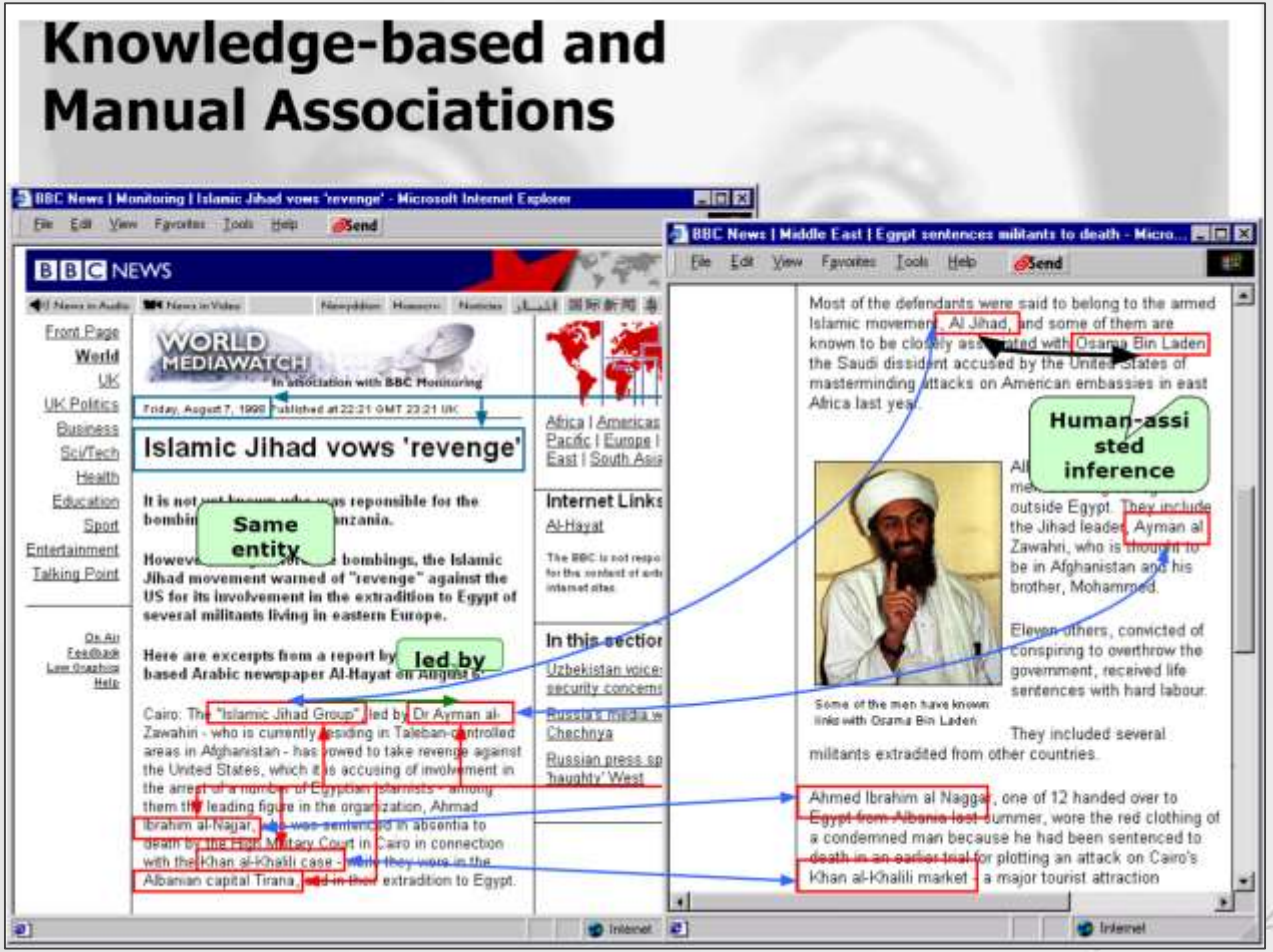


Figure: *Knowledge-based and Manual Associations*

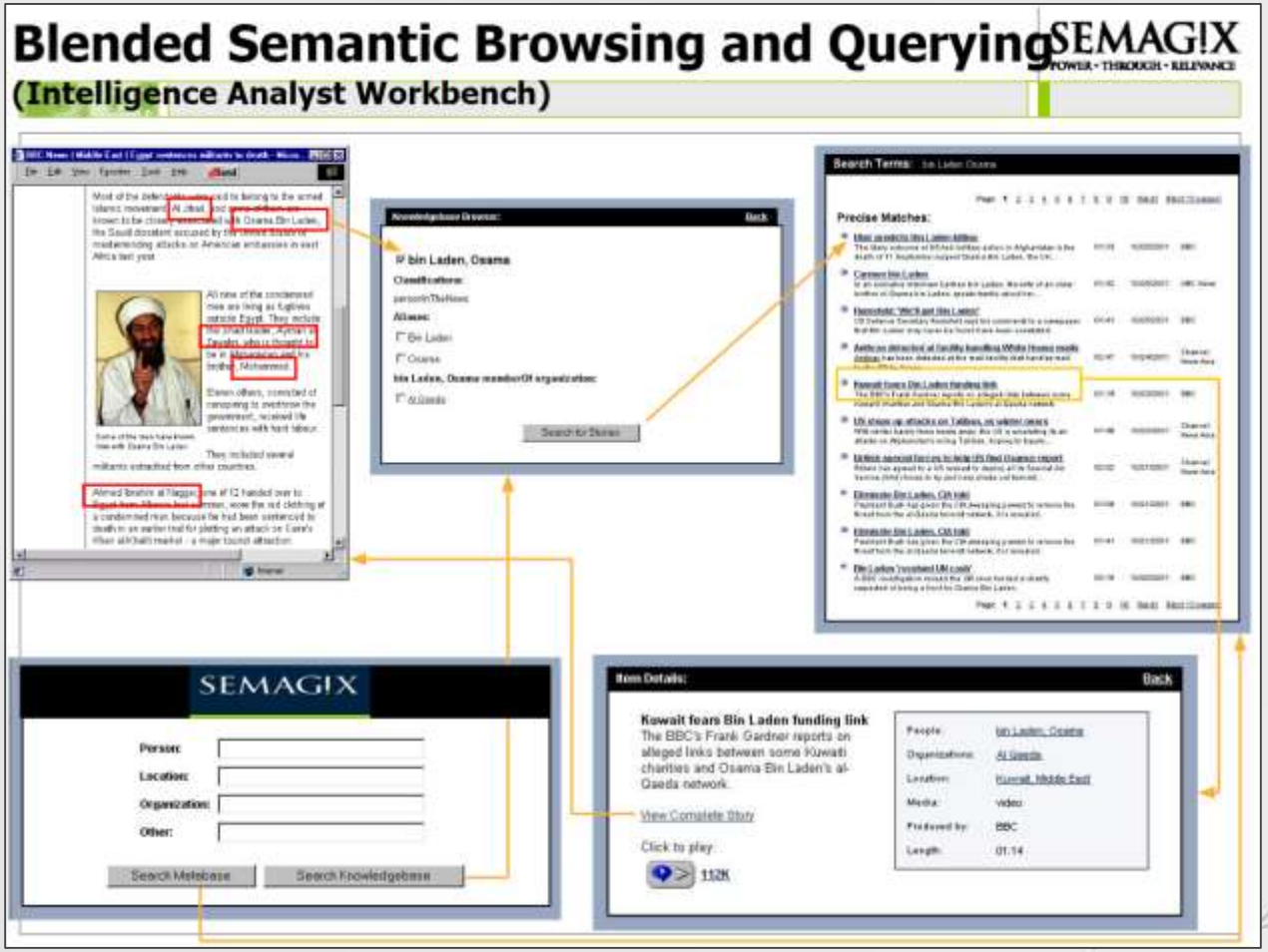


Figure: *Blended Semantic Browsing and Querying*

Semantic EventTracker representation of geospatial and temporal dimensions for semantic associations

- Visualization of association unfolding over time
- Integration of associated multimedia content
- Separate Temporal, Geospatial, and Thematic ontologies describe data
- DEMO

Figure: *Semantic EventTracker*

Semantics for the Semantic Web

The Implicit, the Formal and the Powerful

Knowledge will Propel Machine Understanding of Content: Extrapolating from Current Examples

Amit Sheth
Kno.e.sis Center, Wright State University
Dayton, Ohio, USA
amit@knoesis.org

Sujan Perera
Kno.e.sis Center, Wright State University
Dayton, Ohio, USA
sujan@knoesis.org

Sanjaya Wijeratne
Kno.e.sis Center, Wright State University
Dayton, Ohio, USA
sanjaya@knoesis.org

Krishnaprasad Thirunarayan
Kno.e.sis Center, Wright State University
Dayton, Ohio, USA
tkprasad@knoesis.org

ABSTRACT

Machine Learning has been a big success story during the AI resurgence. One particular stand out success relates to learning from a massive amount of data. In spite of early assertions of the unreasonable effectiveness of data, there is increasing recognition for utilizing knowledge whenever it is available or can be created purposefully. In this paper, we discuss the indispensable role of knowledge for deeper understanding of content where (i) large amounts of training data are unavailable, (ii) the objects to be recognized are complex, (e.g., implicit entities and highly subjective content), and (iii) applications need to use complementary or related data in

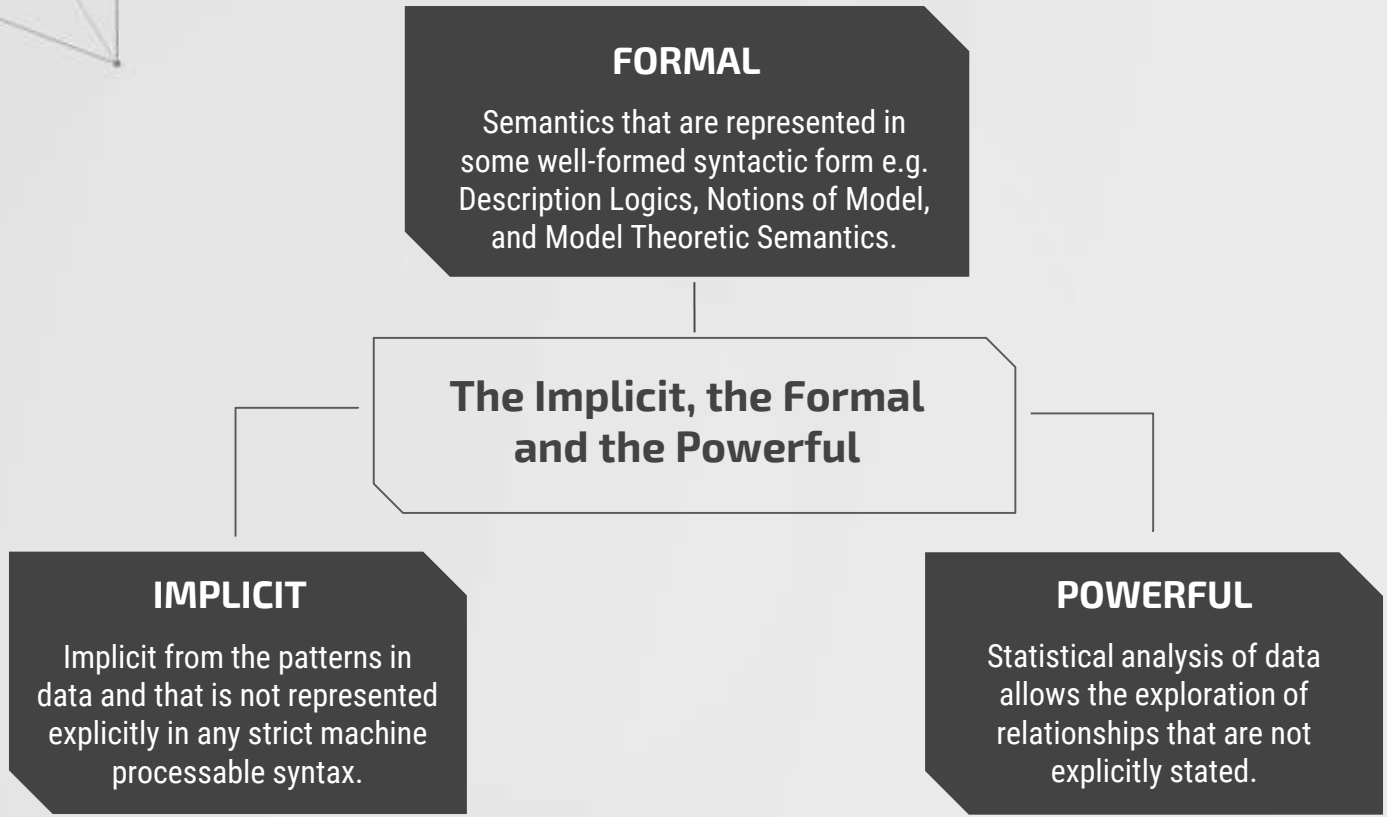
DOI: 10.1145/3106426.3109448

1 INTRODUCTION

Recent success in the area of Machine Learning (ML) for Natural Language Processing (NLP) has been largely credited to the availability of enormous training datasets and computing power to train complex computational models [12]. Complex NLP tasks such as statistical machine translation and speech recognition have

http://knoesis.org/node/2835

Semantics for the Semantic Web



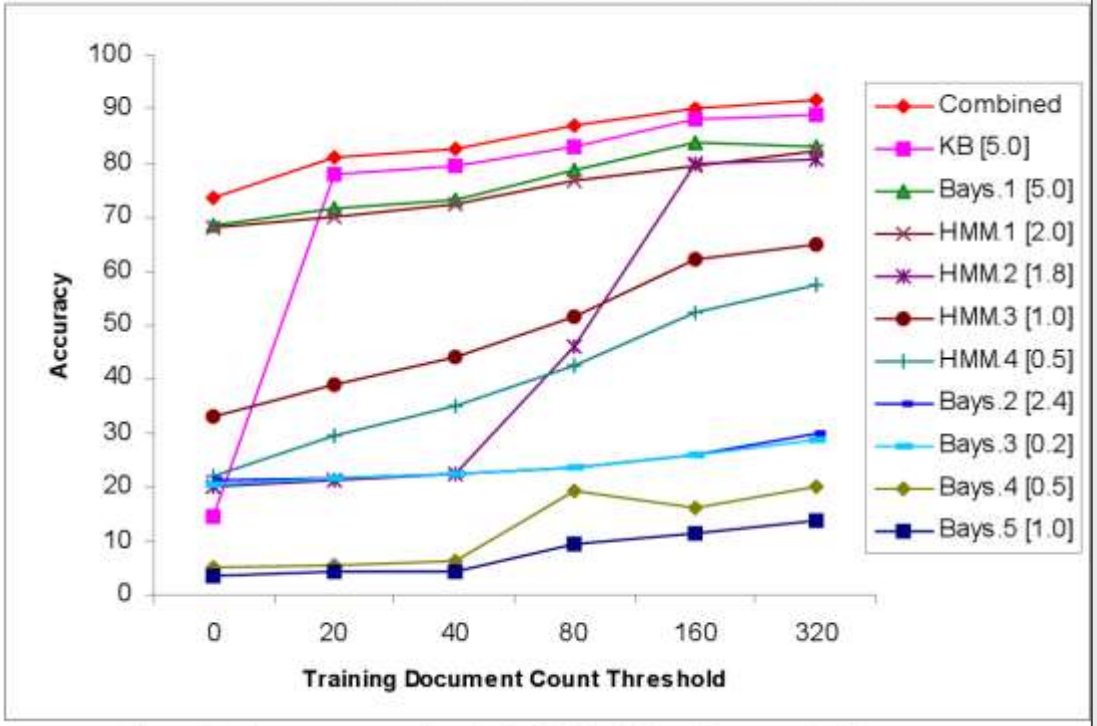
Sheth, A., Ramakrishnan, C., & Thomas, C. (2005). Semantics for the semantic web: The implicit, the formal and the powerful. International Journal on Semantic Web and Information Systems (IJSWIS), 1(1), 1-18.

Semantic Enhancement Engine:

A Modular Document Enhancement Platform for Semantic Applications over Heterogeneous Content

(Brian Hammond, Amit P. Sheth, Krzysztof J. Kochut)

The classification committee consistently outperformed the individual classifiers.



Combining statistical and knowledge-based techniques for NER

Accuracy rates for the individual classifiers on the Reuters set with various threshold values on the number of training documents required.

PAC Learning

Probably Approximately Correct (PAC) ---- Need for Knowledge Infusion

$$samples \geq \frac{\log |C| - \log \delta}{\epsilon}$$

Number of data “samples” required to effectively train a classifier(s) depends on

1. Checking all the possible classifiers (|C|)
2. True error of the data (human annotation/human bias (δ))
3. Generalization error of the model (ε)

This equation make two critical inference:

1. **Confidence:** More Certainty (lower δ) means more number of samples.
2. **Complexity:** More complicated hypothesis (|C|) means more number of samples

https://openreview.net/forum?id=rJxbJeHFPS

How do you know that a training set has a good domain coverage?

A KG (or Ontology) schema is designed by domain experts. It is populated from a representative DB (sets of instances). A KG has very large number of instances (mapping to # of training examples).

How do ensure consistency of labeling, esp when label is not binary?

Do labels represent adequate semantics (e.g., number of alternatives)?

Do they have adequate domain knowledge?

How do you ensure consistency of labeling (interpretation)?

A good KG has addresses these issues:

- a schema is rich in representation (and captures much more than labeling)
- KG design incorporate substantiate domain knowledge
- instance level knowledge is created through (usually) Collective intelligence and curation

Knowledge Will Propel Machine Understanding of Big Data

The indispensable role of knowledge for deeper understanding of content

Sheth, A., Perera, S., Wijeratne, S., & Thirunarayan, K. (2017, August). Knowledge will propel machine understanding of content: extrapolating from current examples. In Proceedings of the International Conference on Web Intelligence (pp. 1-9). ACM.

Knowledge Will Propel Machine Understanding of Big Data

The indispensable role of knowledge for deeper understanding of content where

- Large amounts of training data are unavailable.
- The objects to be recognized are complex (e.g., implicit entities and highly subjective content).
- Applications need to use complementary or related data in multiple modalities/media.

Machine Intelligence - we will interpret it much more broadly than Google: “all aspect of machine learning”... We will define it as machines (any system) performing similar to (nearly emulating) human intelligence.

How will machines “understand” the data/signals/observations, so that it can (help) take timely and good (evidence based) decision and actions.

Knowledge plays an indispensable role in deeper understanding of content

Knowledge often complements or enhances ML and NLP techniques, supports in contextual “understanding” of data.

Especially interesting situations:

I. Large amounts of training data are unavailable,
II. The objects to be recognized are complex, such as implicit entities and highly subjective content, and
III. Applications need to use complementary or related data in multiple modalities/media.

Knowledge-based Approaches
and the Resulting Improvements

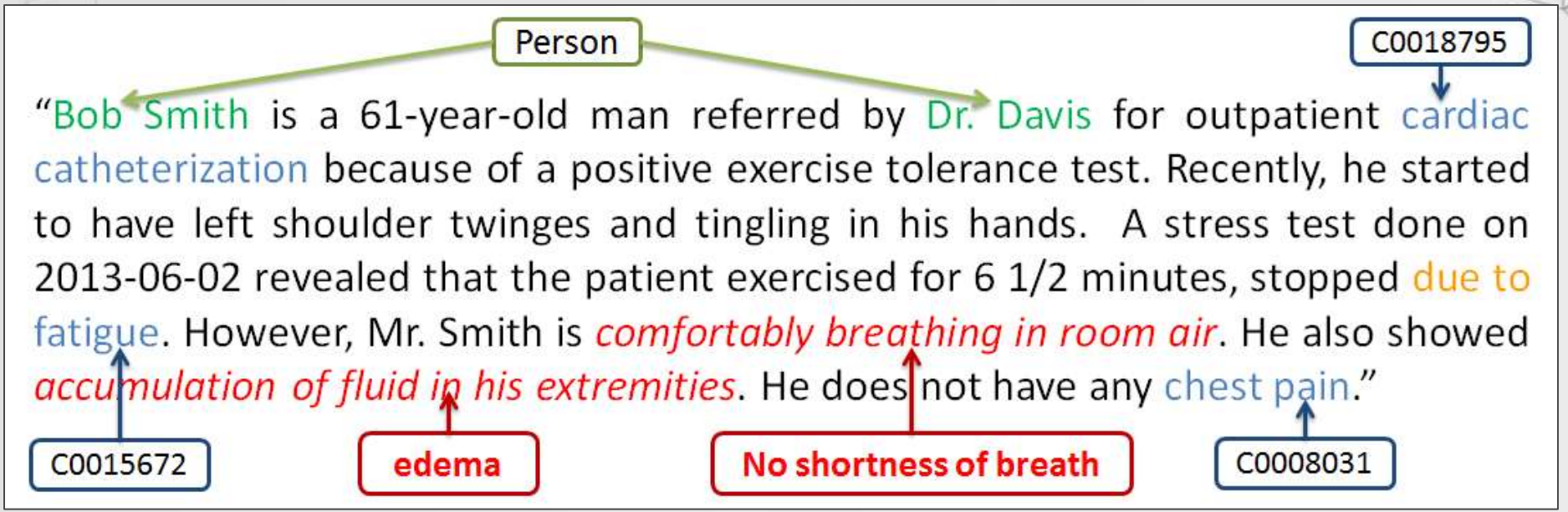
Problem Domain	Use of Knowledge/Knowledge bases	Problems solved that could not be solved (well) w/o knowledge
Implicit Entity recognition and linking	Adapted UMLS definitions for identifying medical entities, and Wikipedia and Twitter data for identifying Twitter entities	Was not solved before
Understanding Drug Abuse-related Discussions on Web forums	Application of Drug Abuse Ontology along with slang term dictionaries and grammar	Not solved well at all
Understanding city traffic dynamics using sensor and textual observations	Statistical knowledge extraction and using ontologies for Twitter event extraction	Multi-modal data stream correlation and explanation virtually impossible
Emoji Similarity and Sense Disambiguation	Generation and application of EmojiNet	Emoji interpretation solved much better

Implicit Entity Recognition and Linking

Sujan Perera, Pablo N. Mendes, Adarsh Alex, Amit Sheth, Krishnaprasad Thirunarayan. Implicit Entity Linking in Tweets. Extended Semantic Web Conference. Heraklion, Crete, Greece : Springer; 2016. p. 118-132. http://knoesis.org/node/2644

Sujan Perera, Pablo Mendes, Amit Sheth, Krishnaprasad Thirunarayan, Adarsh Alex, Christopher Heid, Greg Mott. Implicit Entity Recognition in Clinical Documents. 4th Joint Conference on Lexical and Computational Semantics (*SEM) 2015. Denver, CO: Association for Computational Linguistics; 2015. p. 228-238. http://knoesis.org/node/2171

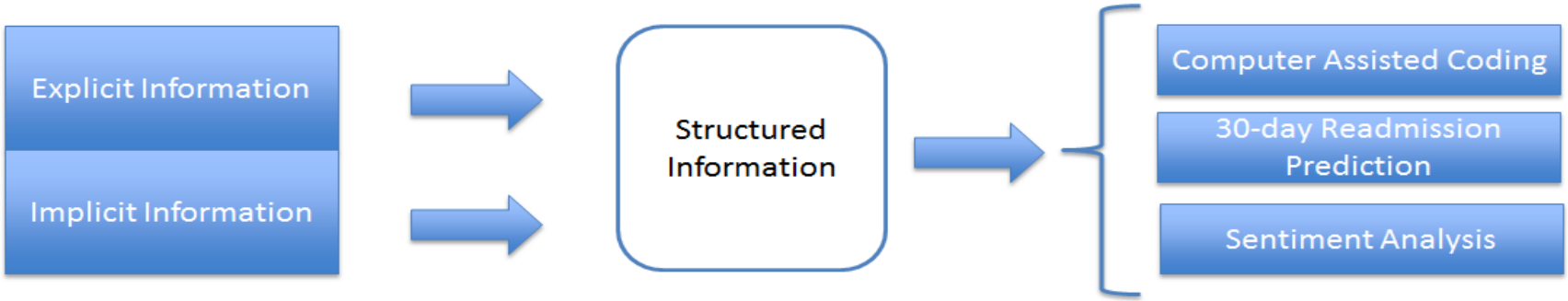
Implicit Entity Recognition and Linking



Named Entity Recognition	Relationship Extraction	Entity Linking	Implicit information extraction

Significance

- Volume
  - 20% movie references and 40% book references in tweets
  - 35% edema and 40% shortness of breath references in clinical narratives
- Value



Ignoring implicit information in text would adversely affect downstream applications

Role of Knowledge

The patient showed accumulation of fluid in his extremities, but respirations were unlabored and there were no use of accessory muscles.

Knowledge

Edema	Accumulation of an excessive amount of watery fluid in cells or intercellular tissues
Shortness of breath	Labored or difficult breathing associated with a variety of disorders

New Sandra Bullock astronaut lost in space movie looks absolutely terrifying

Sandra Bullock → Gravity

	Entity Type	Without Contextual Knowledge	With Contextual Knowledge
Disambiguation Accuracy	Movie	51.7%	**60.97%**
	Book	50.0%	**61.05%**

Understanding and Analyzing Drug Abuse Related Discussions on Web Forums

Cameron, Delroy, Gary A. Smith, Raminta Daniulaityte, Amit P. Sheth, Drashti Dave, Lu Chen, Gaurish Anand, Robert Carlson, Kera Z. Watkins, and Russel Falck. "PREDOSE: a semantic web platform for drug abuse epidemiology using social media." Journal of biomedical informatics 46, no. 6 (2013): 985-997. http://knoesis.org/node/2469

Knowledge Bases:
Application of Drug Abuse Ontology along with slang term dictionaries and grammar

Nature of Improvement:
Recall and coverage

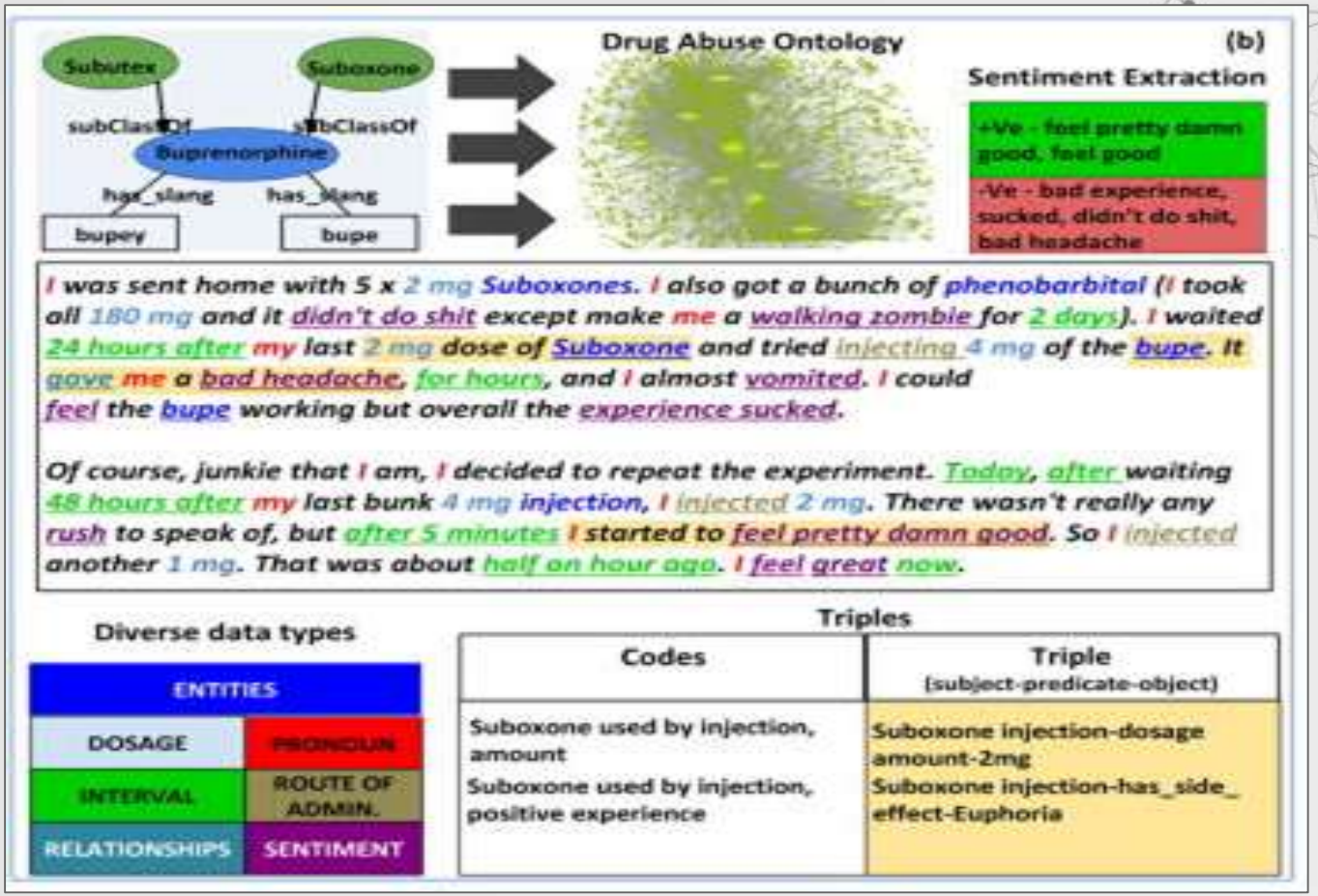
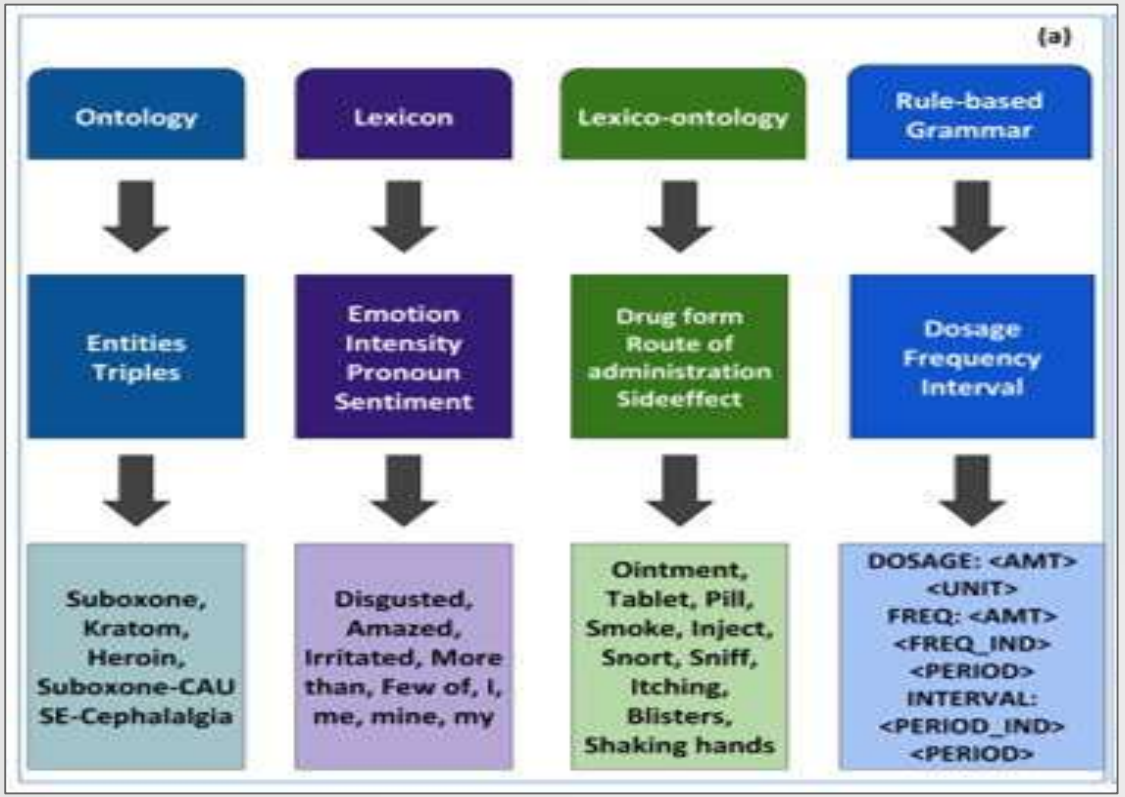


Figure: Understanding and analyzing drug abuse related discussions on web forums

PREDOSE:

Smarter Data through Shared Context and Data Integration



Understanding city traffic using sensor and textual observations

Pramod Anantharam, Krishnaprasad Thirunarayan, Surendra Marupudi, Amit Sheth, Tanvi Banerjee. Understanding City Traffic Dynamics Utilizing Sensor and Textual Observations. In 30th AAAI Conference on Artificial Intelligence (AAAI-16). Phoenix, Arizona; 2016. http://knoesis.org/node/2145

Pramod Anantharam, Krishnaprasad Thirunarayan, Amit Sheth. Traffic Analytics using Probabilistic Graphical Models Enhanced with Knowledge Bases. In 2nd International Workshop on Analytics for Cyber-Physical Systems (ACS-2013) at SIAM International Conference on Data Mining (SDM 2013). Austin, Texas; 2013. http://knoesis.org/node/2476

Severity of the Traffic Problem

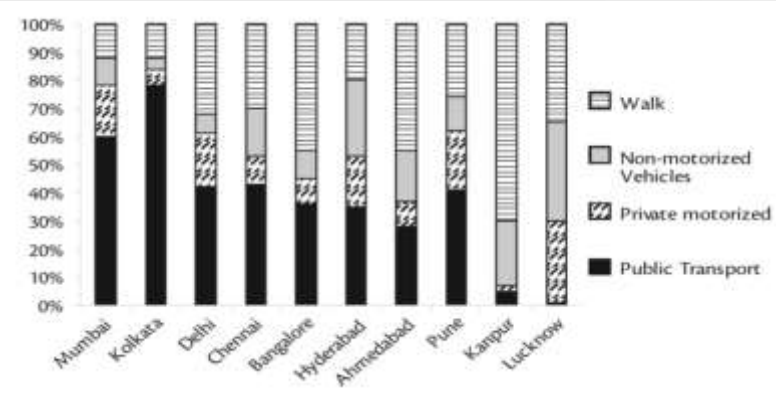
By 2001 over 285 million Indians lived in cities, more than in all North American cities combined (Office of the Registrar General of India 2001) [1].

[1] The Crisis of Public Transport in India.
[2] IBM Smarter Traffic.

Today there are more than [2011]
one billion cars on the road.
That number **will double by 2030**.

Traffic jumped **236%** as population grew nearly 20% between 1982 and 2001 in the U.S.

Modes of Transportation in Indian Cities



The Texas Transportation Institute (TTI) Congestion report for the United States

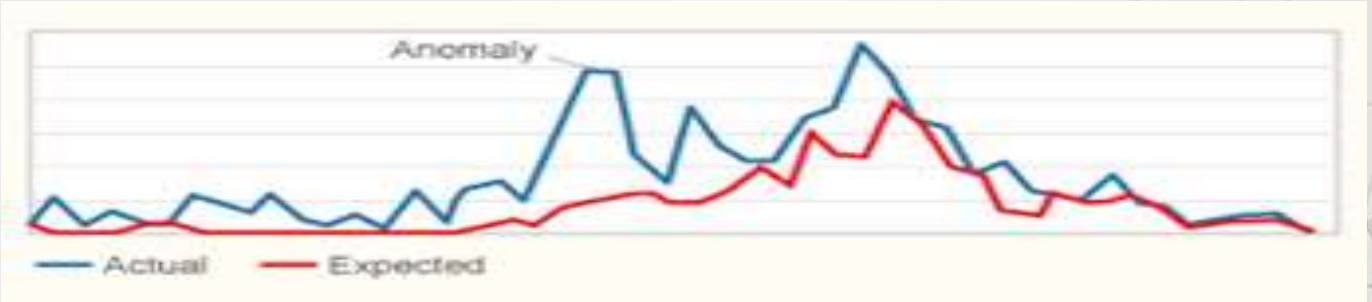
The TTI report analyzed the impact of public transportation in 439 metropolitan areas, categorized as very large, large, medium and small.

- Very large areas (3+ million): Public transportation saved 557 million hours of delay and $11.9 billion in congestion costs.
- Large areas (1-3 million): 59 million hours of delay and $1.2 billion saved.
- Medium urban areas (500,000-1 million): 13 million hours and $259 million saved.
- Small areas (less than 500,000): Public transportation saved 2 million hours of delay and $31 million.

Questions Asked Daily

- What time to start?
- What route to take?
- What is the reason for traffic?
- Wait for some time or re-route?

Complementary Data Sources



Challenge: Non-linearity in Traffic Dynamics

Multiple Events

Interact with Each Other

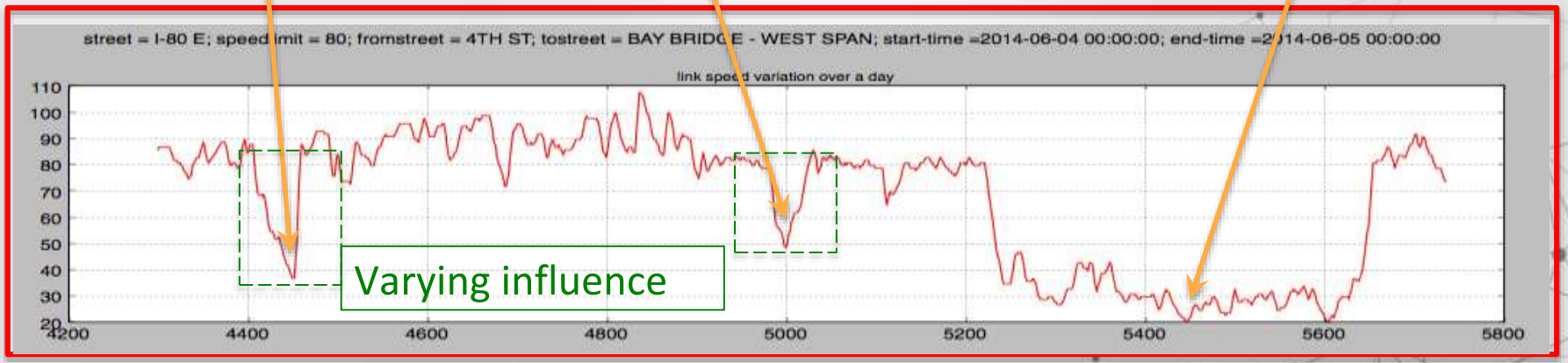


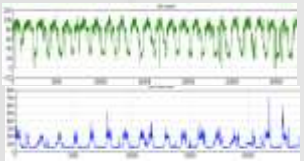
Image credit: http://traffic.511.org/index.

Learning Context-specific LDS Models

Step 1: Index data for **each link** for **day of week** and **hour of day** utilizing the traffic domain knowledge for piece-wise linear approximation

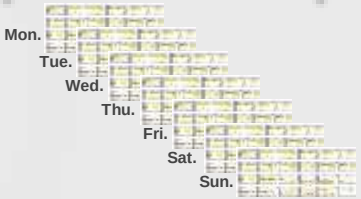
Step 2: Find the "typical" dynamics by computing the *mean* and choosing the *medoid* for each hour of day and day of week

Step 3: Learn LDS parameters for the *medoid* for each hour of day (24 hours) and each day of week (7 days) resulting in 24 × 7 = 168 models for each link



Speed/travel-time time series data from a link.

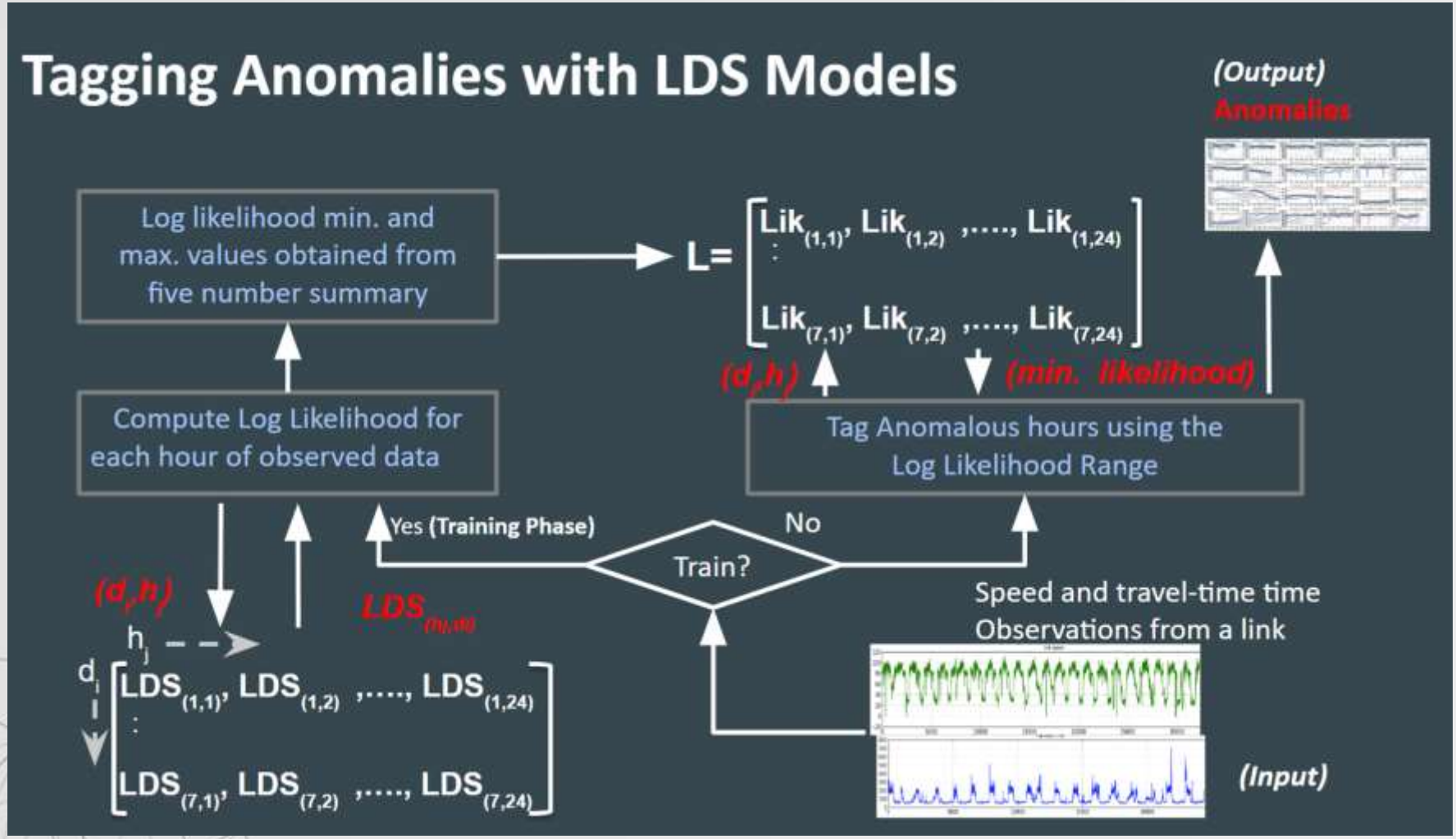
Time series data for each hour of day (1-24) for each day of week (Monday – Sunday).



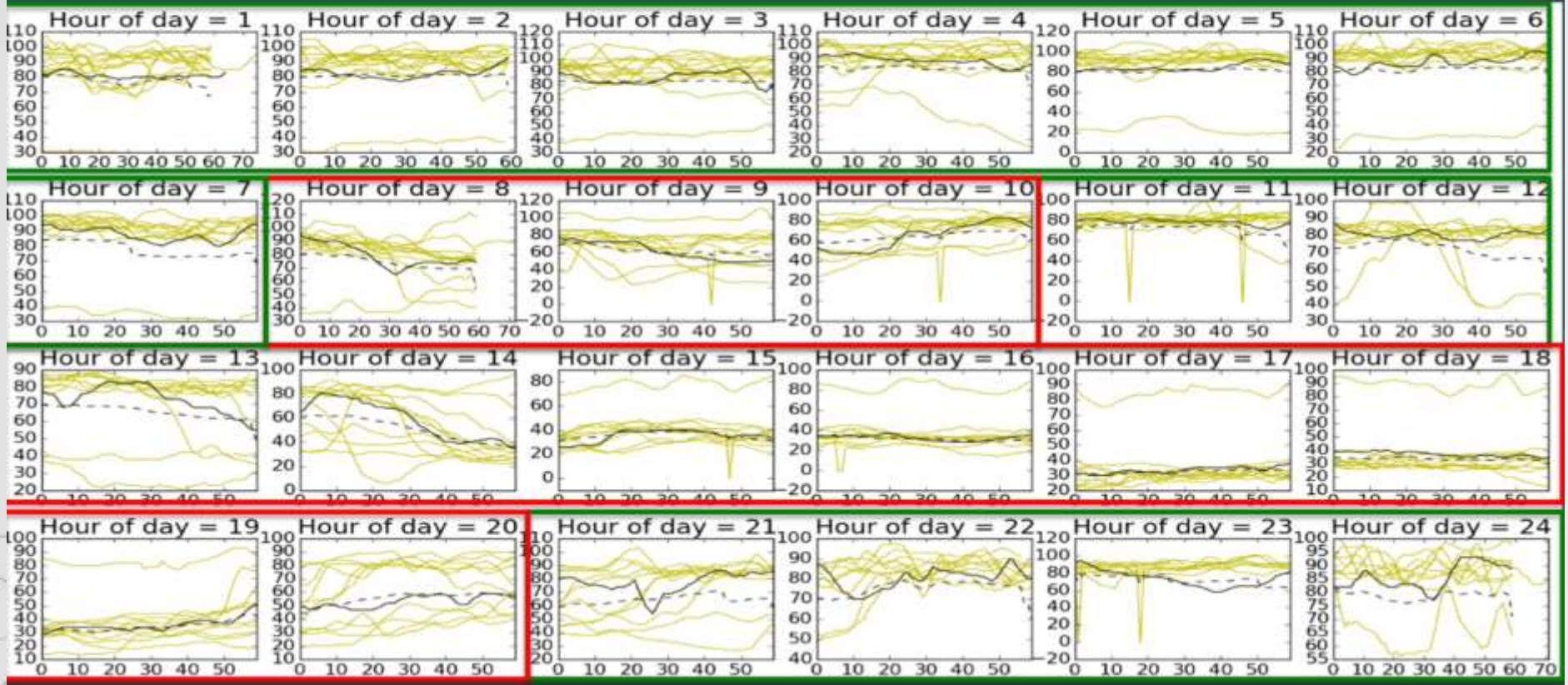
Mean time series computed for each day of week and hour of day along with the medoid.

168 LDS models for each link; Total models learned = **425,712** i.e., (2,534 links × 168 models per link).

Tagging Anomalies with LDS Models

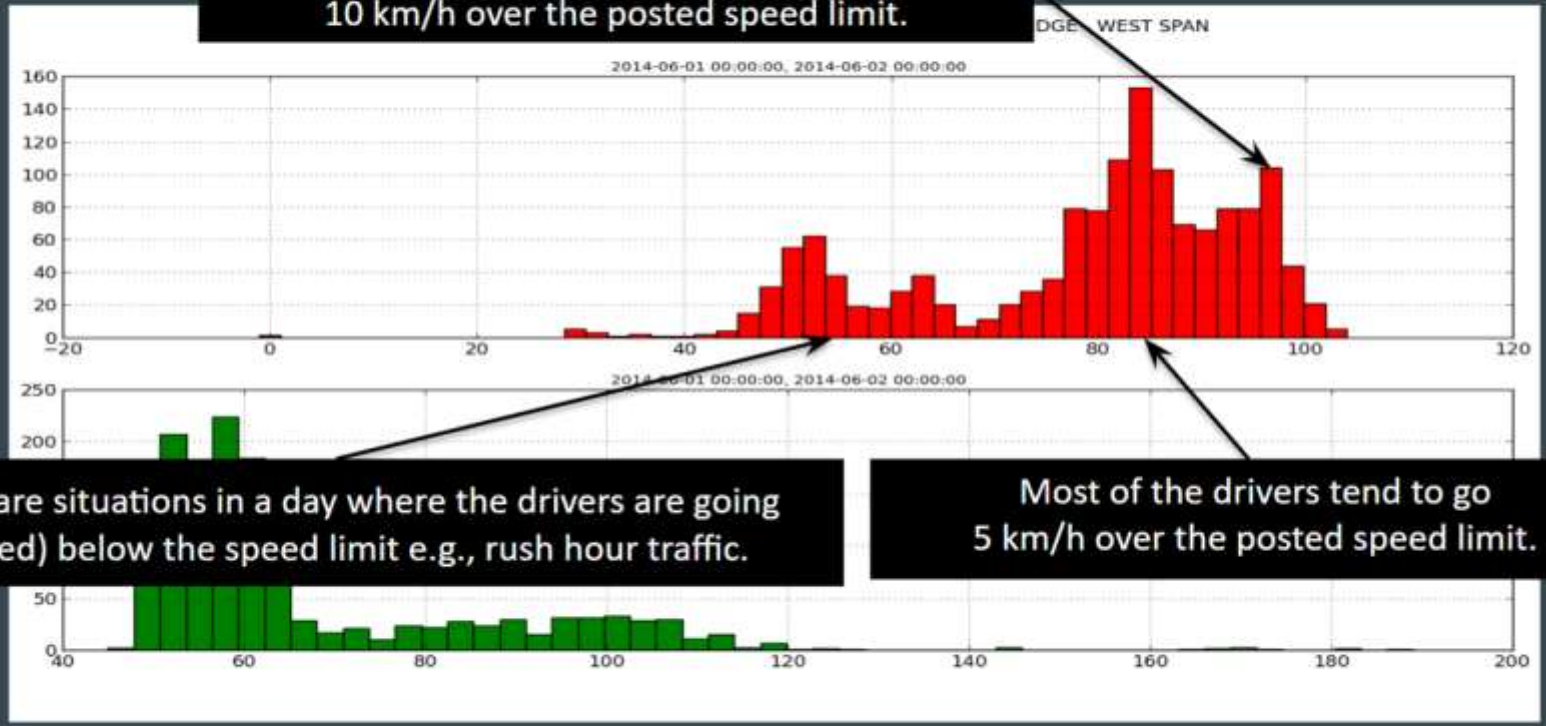


Hourly Traffic Dynamics Over a Day



Traffic Data: Possible Explanation

There are relatively few drivers who go more than 10 km/h over the posted speed limit.



There are situations in a day where the drivers are going (forced) below the speed limit e.g., rush hour traffic.

Most of the drivers tend to go 5 km/h over the posted speed limit.

Do these histograms resemble any probability distribution?

Twitter as a Source of City Events

Public Safety

Post Local @postlocal 10 Sep
Small explosion outside Montgomery County's **public safety** building, possibly a firecracker wapo.st/19ExveU
View summary · Reply · Retweet · Favorite · More

Urban Planning

Kansas City Scout @KansasCityScout 8m
My KC Scout Traffic Alert Cleared: **Roadwork** on I-435 NB (East of KC) PAST 24 HWY
Message #65072-142
Expand · Reply · Retweet · Favorite · More

Gov. & Agency Admin.

Oliver Galak @OliverGalak 4 Jun
"AFA, whose president is a firm ally of the Kirchner **administration**", dice el aliento fétido de ABC news abcnews.go.com/ABC_Univision/... #LarrataMoore
View summary · Reply · Retweet · Favorite · More

Social Programs

Elle A from Elle A @elle_emm_aitch 19m
hungry w/ no hope & no **social programs** to access if you suffer from hardship. They want you terminally ill w/o healthcare.
Expand · Reply · Retweet · Favorite · More

Healthcare

Jane Frawley @JaneFrawley 1m
Australian Government urged to invest more in complementary **healthcare** foodnavigator-asia.com/Policy/New-Aus... via @FoodNavAsia
Expand · Reply · Retweet · Favorite · More

Education

Greg Gorman @gormang 1h
Being a Courageous Educator and School Leader in an Era of Foolish **Education Policy**
the21stcenturyprincipal.blogspot.com/2013/09/being-... via @Digg
Expand · Reply · Retweet · Favorite · More

Energy & water

Alan Hawkins @ECjourno 5 Sep
City Power workers turn off the lights in parts of #Johannesburg because of a new shift system #strikeseason news24.com/SouthAfrica/Ne...
View summary · Reply · Retweet · Favorite · More

Kim Hernandez @Kimm31 16 Sep
@unitedutilities please please please will you fix the air in the **water supply problem**,CH46, my kids refuse to drink the water
Expand · Reply · Retweet · Favorite · More

Environmental

elaman Copp @HelamanCopp 13 Sep
e're at Muthurwana Secondary School in Sundani Village, **Waste anagement** is a **problem**, for one, uQshdgbr7k
Expand · Reply · Retweet · Favorite · More

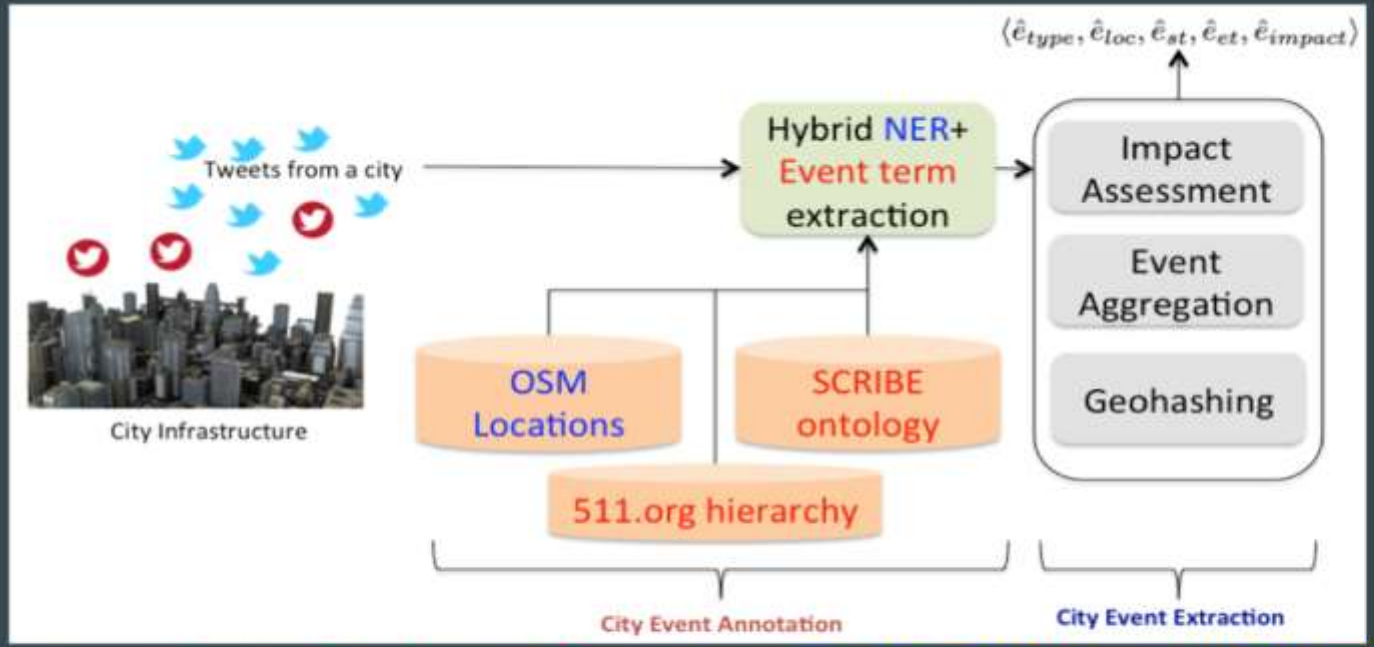
Autogas Limited @AutogasLPG 31 Aug
#Warwick resident? Poor air quality is a **problem** in the town. Switching to LPG could help tackle **air pollution** ow.ly/oIRrs
pand · Reply · Retweet · Favorite · More

Transportation

Ram N Kumar @ramnkumar 51s
Crossing Yamuna now, it is mirred with **traffic Jam** and hooliganism by people going for Ganpati idol immersion at... fb.me/2IZjrGDOi
Expand · Reply · Retweet · Favorite · More

Erikaaaa @erikaabethl97 2h
@louisacc im at the station now! My **bus** came ahha beware there are loads of **delays** at the station
View conversation · Reply · Retweet · Favorite · More

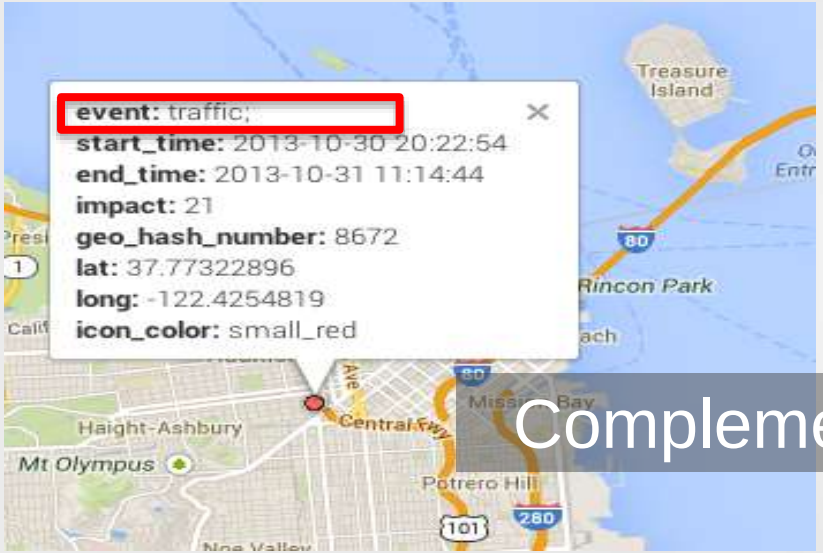
Extracting City Events from Textual Data



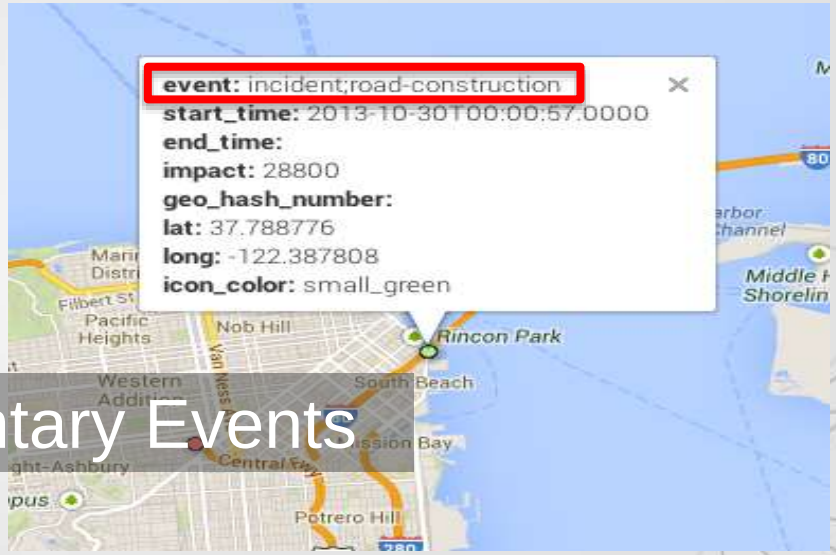
Last O night O in O CA... O (@ O Half B-LOCATION Moon I-LOCATION Bay B-LOCATION Brewing I-LOCATION Company O w/ O 8 O others) O http://t.co/w0eGEJjApY O

Pramod Anantharam, Payam Barnaghi, Krishnaprasad Thirunarayan, and Amit Sheth. 2015. **Extracting City Traffic Events from Social Streams. ACM Trans. Intell. Syst. Technol.** 6, 4, Article 43 (July 2015), 27 pages. DOI: 10.1145/2717317. http://doi.acm.org/10.1145/2717317/

Textual Events from Tweets vs. 511.org: Complementary



Traffic



Incident;
road-construction

Complementary Events

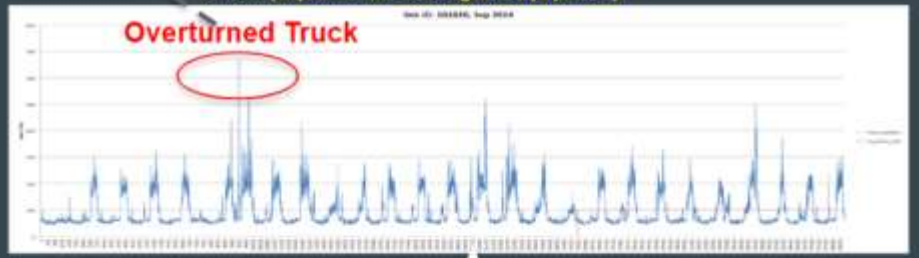
Understanding: Semantic Annotation of Sensor + Textual Data Utilizing Background Knowledge

Domain knowledge in the form of traffic vocabulary

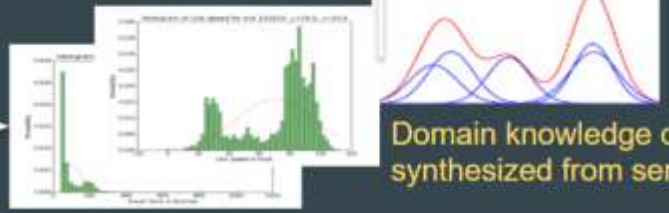
Accident B-EVENT on O the O Golden B-LOCATION Highway I-LOCATION at O the O Viking O robots O in O Devland O JHB O , O ambo O truck O , O injured B-EVENT treating O themselves O

Explained-by

Horizontal operator: relating/mapping data from different modality to a concept (theme) within a spatio-temporal context;
Spatial context even include what it means to have a slow *traffic for the type of road (http://wiki.knoesis.org/index.php/PCS)*



511



Domain knowledge of traffic flow synthesized from sensor data

Image Credit:
http://traffic.511.org/index

How traffic analysis captures complexity of the real-world?

This example demonstrates use of:

- Multimodal data streams (types of events from text - signature from sensor data).
- Multiple sources of declarative knowledge/ontologies.
- Semantic annotations and enrichments.
- Use of rich representation (PGM)
  - learned probabilistic models improved using declarative knowledge

- Statistical approach to create normalcy models and understand anomalies using historical data. Explain anomalies using extracted events.
  - use declarative knowledge to approximate nonlinear models using a collection of linear dynamical systems

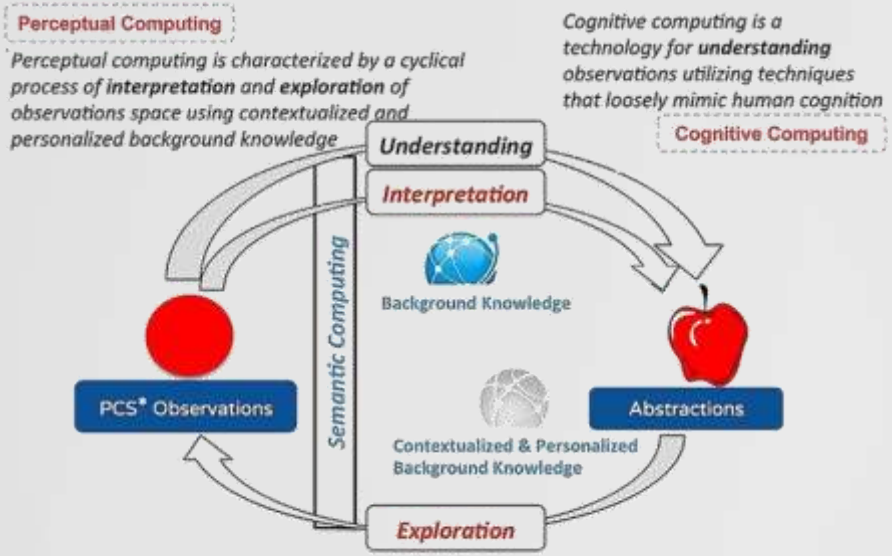
- Provide actionable information.

Semantic, Cognitive, and Perceptual Computing

Paradigms That Shape Human Experience

Semantic, Cognitive, and Perceptual Computing:

Paradigms That Shape Human Experience



http://bit.ly/SCPComputing

Humans are interested in **high-level concepts** (phenotypic characteristics).

Semantic Computing: Assign labels and associate meanings (representation & contextualization).

Cognitive Computing: Interpretation of data with respect to perspectives, constraints, domain knowledge, and personal context.

Perceptual Computing: A cyclical process of semantic-cognitive computing for higher level of perception and reasoning (abstraction & action).

Sheth, A., Anantharam, P., & Henson, C. (2016). Semantic, cognitive, and perceptual computing: Paradigms that shape human experience. Computer, 49(3), 64-72.

Interplay between
Semantic, Cognitive, and Perceptual Computing

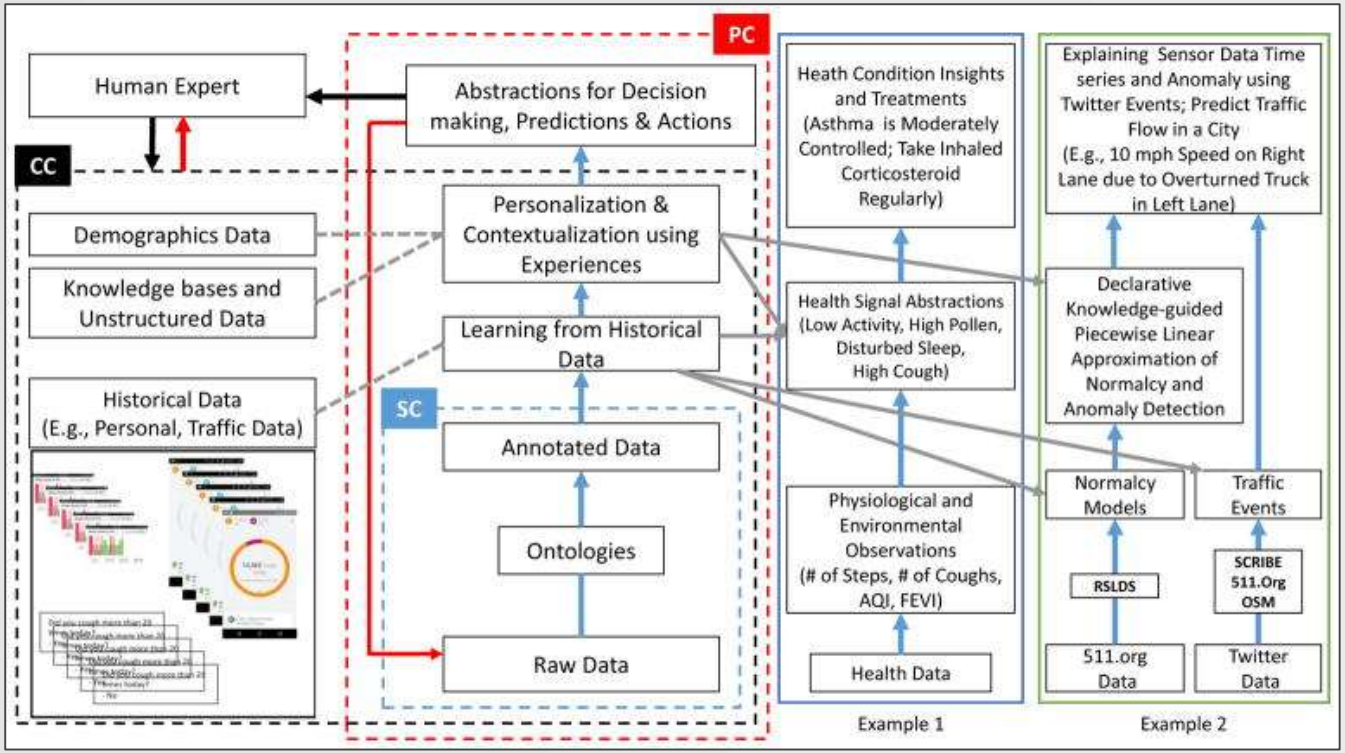


Figure: *SCP with Examples on Asthma and Traffic Use-cases*

PRESENT

Knowledge Graph (Infusion) for Deep Learning

BIG DATA

Top-down symbolic approach (concepts, rules) in data reasoning, inferencing, and deduction.

SEMANTIC & KNOWLEDGE GRAPH

Bottom-up statistical approach in searching, analyzing and deriving insights from Big Data.

MACHINE/ DEEP LEARNING

Deep Learning Critical Review

- **Deep Learning thus far is data hungry :** Human beings can learn abstract relationships in a few trials (explicit and implicit definitions).

- **Deep Learning thus far is shallow and has limited capacity for transfer :** Each model is very specific for the task and generalization is not straightforward.

- **Deep Learning thus far has no natural way to deal with hierarchical structure :** Hierarchical structure in English sentences aren't captured in sequential techniques (which motivate move from Sequential LSTMs to Tree LSTMs).

https://arxiv.org/pdf/1503.00075.pdf

Deep Learning Critical Review (Continued...)

- **Deep Learning thus far has struggled with open-ended inference:** Current machine reading systems have achieved some degree of success in tasks like **SQuAD**, in *which the answer to a given question is explicitly contained within a text, but far less success in tasks in which inference goes beyond what is explicit in a text, either by combining multiple sentences (so called multi-hop inference)* or by combining explicit sentences with background knowledge that is not stated in a specific text selection.

- **Deep Learning thus far has not been well integrated with prior knowledge:** Consider a set of easily-drawn inferences that people can readily answer without anything like direct training, such as *Who is taller, Prince William or his baby son Prince George? Can you make a salad out of a polyester shirt? If you stick a pin into a carrot, does it make a hole in the carrot or in the pin?* As far as I know, nobody has even tried to tackle this sort of thing with deep learning. Such apparently simple problems require humans to integrate knowledge across vastly disparate sources, and as such are a long way from the sweet spot of deep learning-style perceptual classification.

WHY
KNOWLEDGE-INFUSED LEARNING?

- **Ambiguous** online healthcare communications and **difficult** to engineer discriminative features.
- Domain-specific embedding models provide a **shallow infusion** of knowledge.
- Decrease the **dependence** on large datasets
- **Reduce bias** in the dataset (ie: potentially avoid social discrimination and unfair treatment)
- Provide **information provenance**: Allowing explainability of a model
- **Improve information coverage** specific to a domain that would be missed otherwise
- **Reduce time and space complexity** of the models architecture
- Improve models **sensitivity** and **specificity**

KNOWLEDGE-INFUSED LEARNING

SHALLOW Infusion

of knowledge graphs to improve the semantic and conceptual processing of data.

Deeper and congruent incorporation or integration of the knowledge graphs in the learning techniques.

SEMI-DEEP Infusion

DEEP Infusion

(Part of Future KG Strategy)

combines statistical AI (bottom-up) and symbolic AI learning techniques (top-down) for hybrid and integrated intelligent systems.

Shallow Infusion

In shallow infusion, both the external information and method of knowledge infusion is shallow.

Shallow External Knowledge

- Shallow external knowledge is described as those form of information which are extracted from text based on some heuristics, often designed for task-specific problems:
  - Bag of Words/Phrases from Corpus
  - Bag of Words/Phrases from Semantic Lexicons
  - Count of Nouns, Pronouns, Verbs
  - Sentiment and Emotions of the sentence
  - Latent topics describing the documents
  - Label assignment to words or phrases in sentence:
    - Mary sold the book to John
    - Mary (agent), the book (theme), John (recipient), sold (predicate)

Shallow Method of Knowledge Infusion

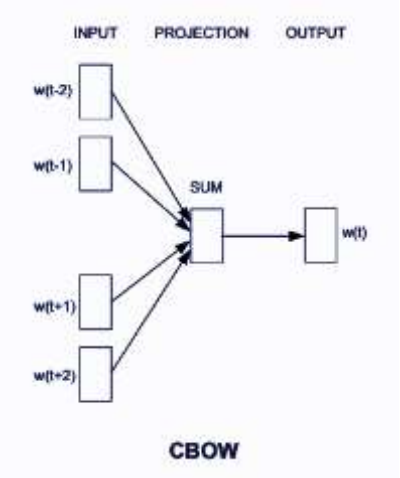
- **Statistical NLP Methods:**
  - Term Frequency and Inverse Document Frequency
  - Latent Dirichlet Allocation/ Hierarchical Dirichlet Process
  - Semantic Role Labeling
  - Zipf Law

- **Neural NLP Methods:**
  - Word2Vec, GLoVE
  - Sentence2Vec, Doc2Vec
  - BERT, ELMo

Examples of Shallow Infusion from NLP domain

Word2Vec

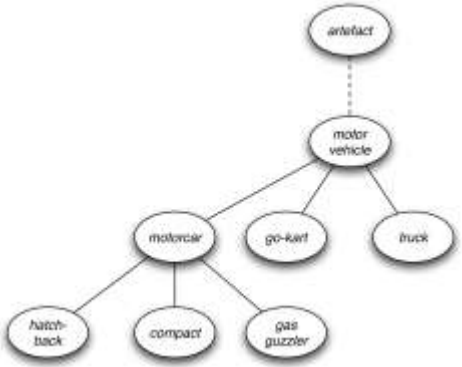
Knowledge: Domain specific large corpora



"Context is represented by a set of words for a given target word"

Retrofitting

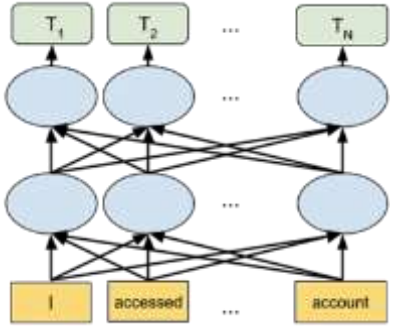
Knowledge: pre-trained embeddings + semantic lexicons



*"Learned embeddings are further **enriched** by using semantic lexicons"*

BERT

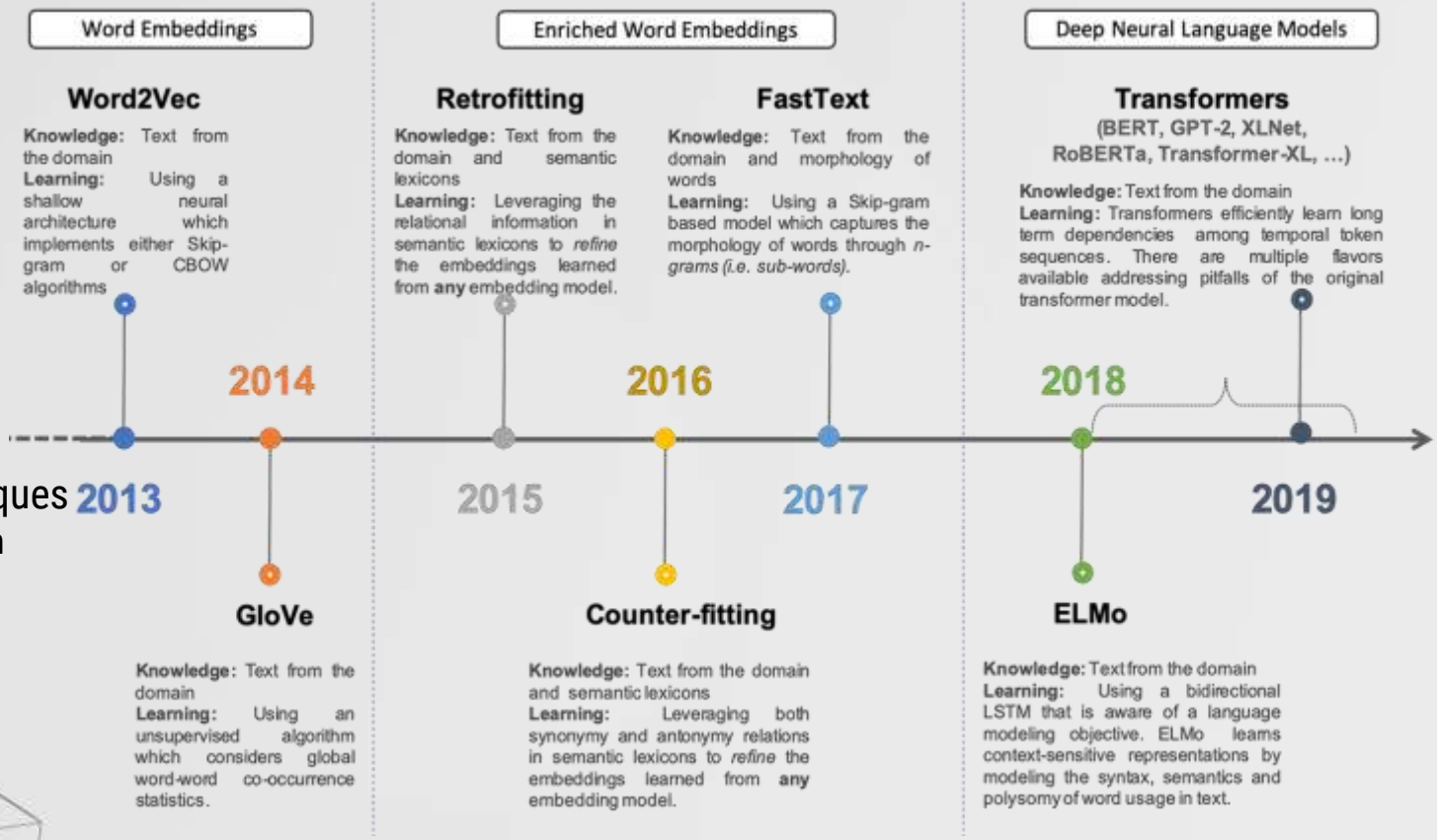
Knowledge: Domain specific large corpora

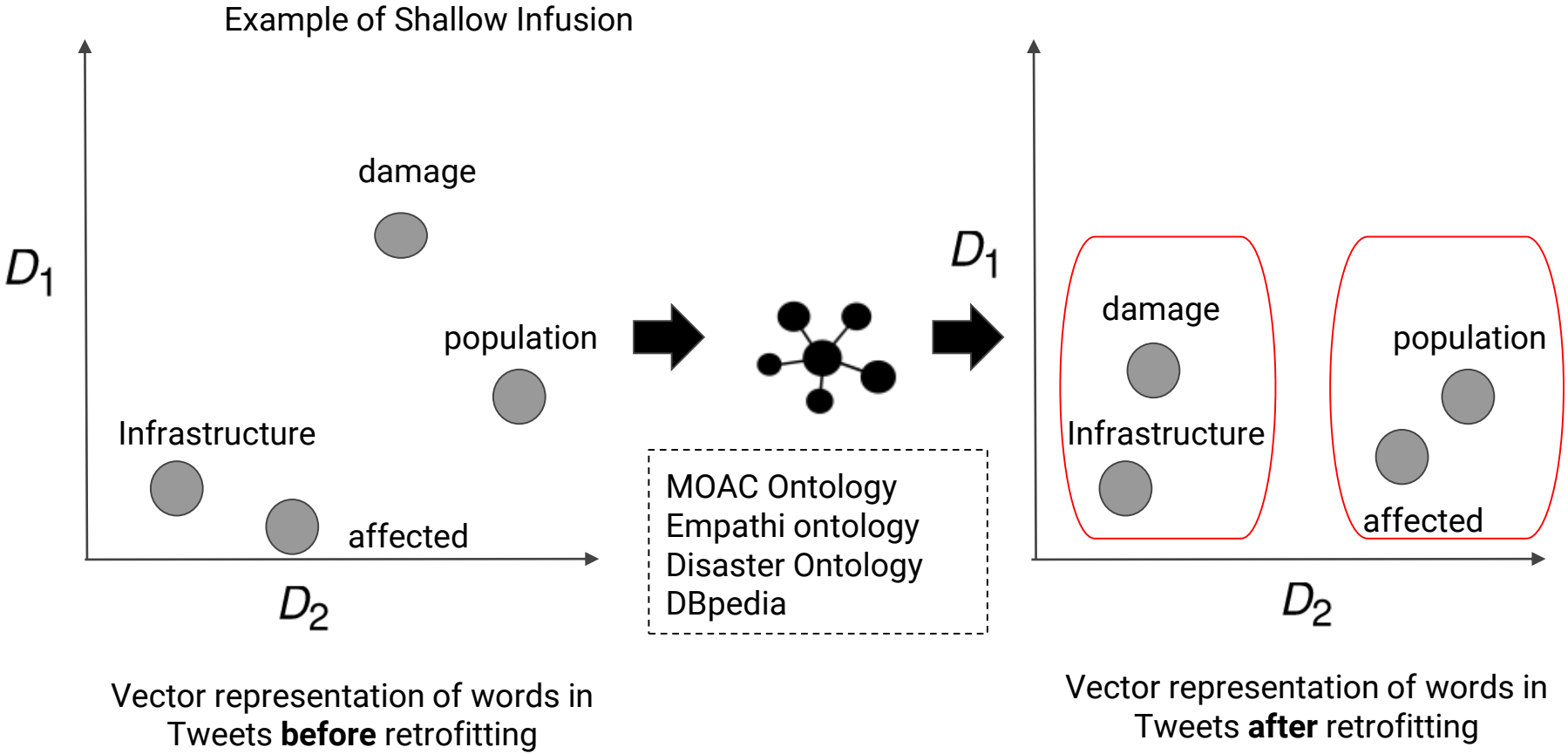


"Uses language modeling objective to learn the contextual representations"

SHALLOW Infusion of Knowledge for Machine/ Deep Learning in Brief

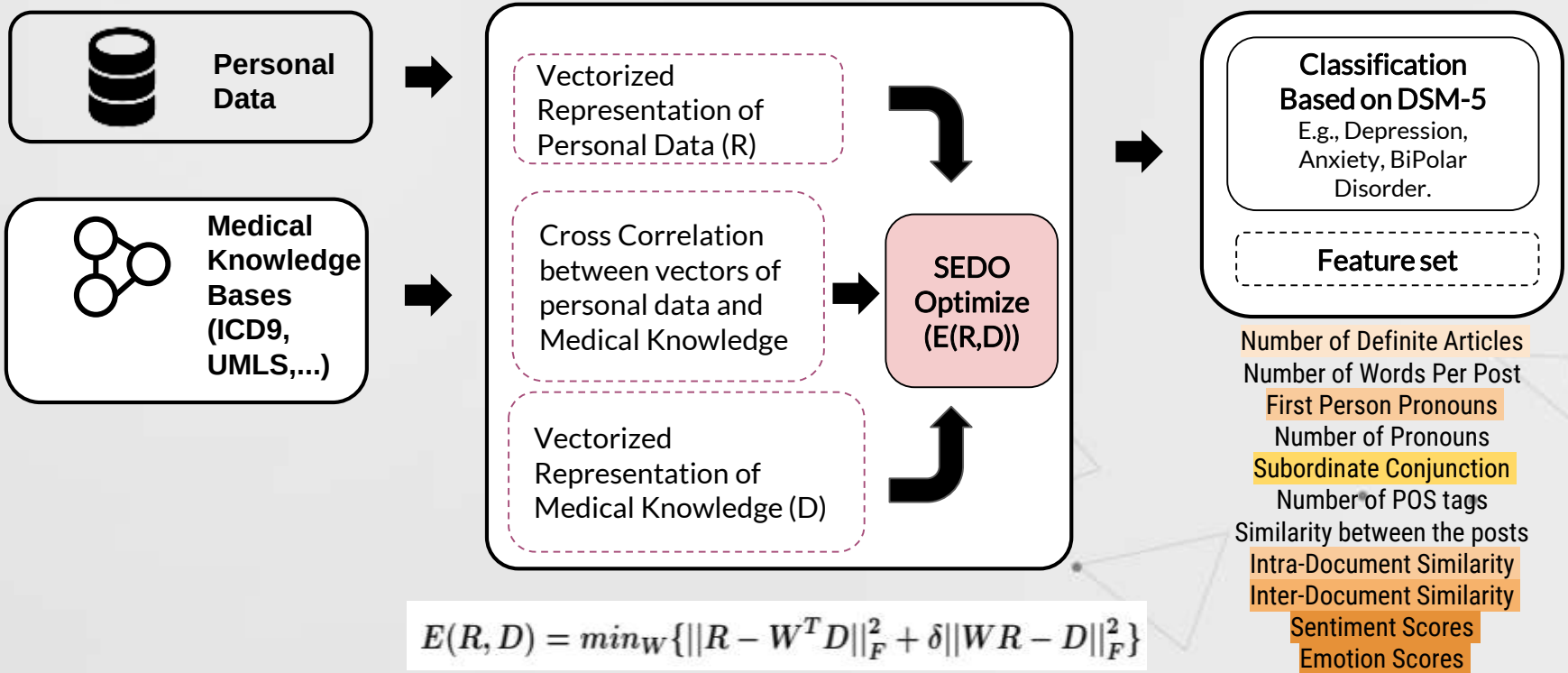
Chronological arrangement of shallow Infusion techniques From NLP domain





Gaur, Manas, et al. "empathi: An ontology for Emergency Managing and Planning about Hazard Crisis." *2019 IEEE 13th International Conference on Semantic Computing (ICSC)*. IEEE, 2019.

Explaining the prediction of mental health disorders (CIKM 2018)

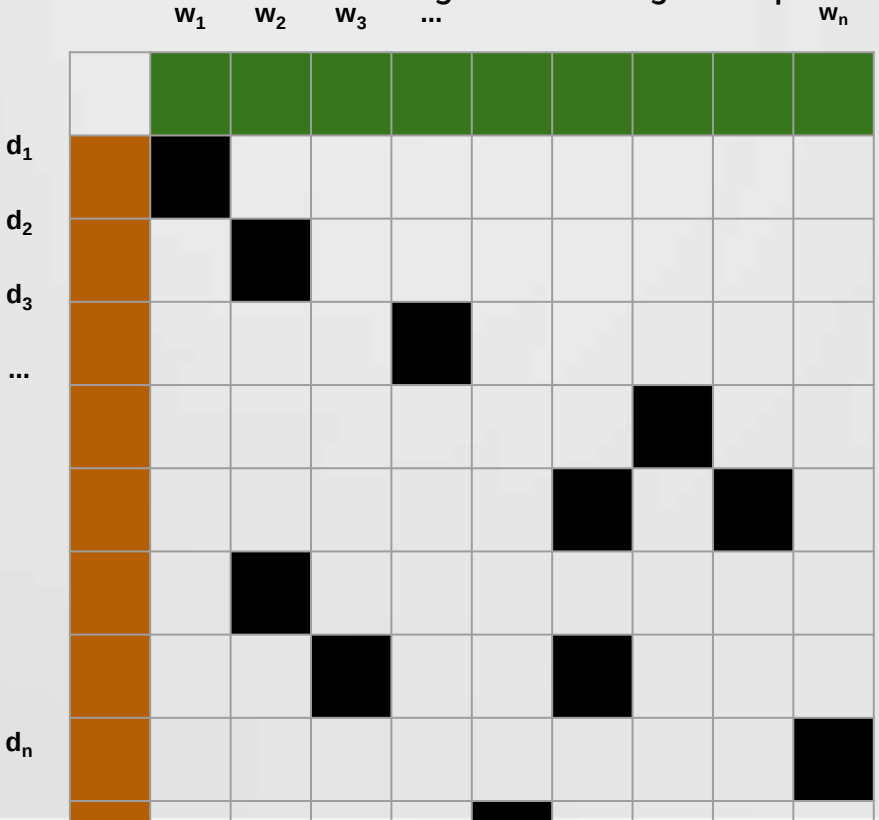


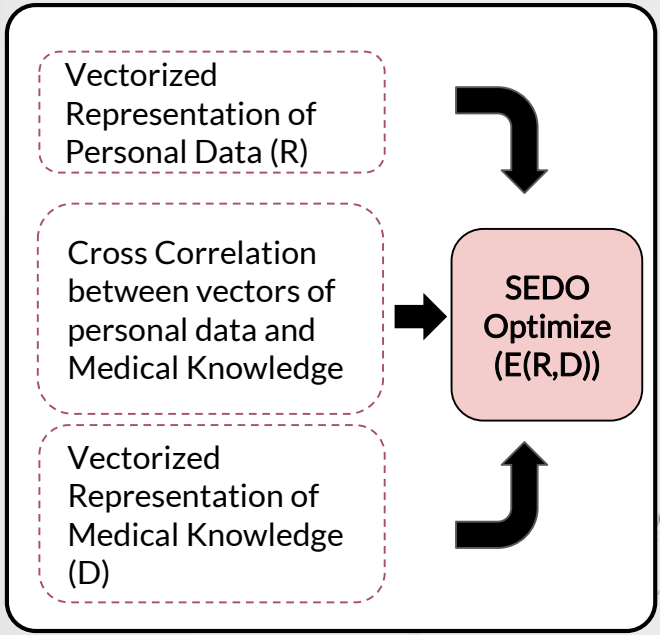
$$E(R,D) = min_W\{||R - W^T D||_F^2 + \delta||WR - D||_F^2\}$$

Manas Gaur, Ugur Kursuncu, Amanuel Alambo, Amit Sheth, Raminta Daniulaityte, Krishnaprasad Thirunarayan, Jyotishman Pathak. "Let Me Tell You About Your Mental Health!" Contextualized Classification of Reddit Posts to DSM-5 for Web-based Intervention. In The 27th ACM CIKM 2018

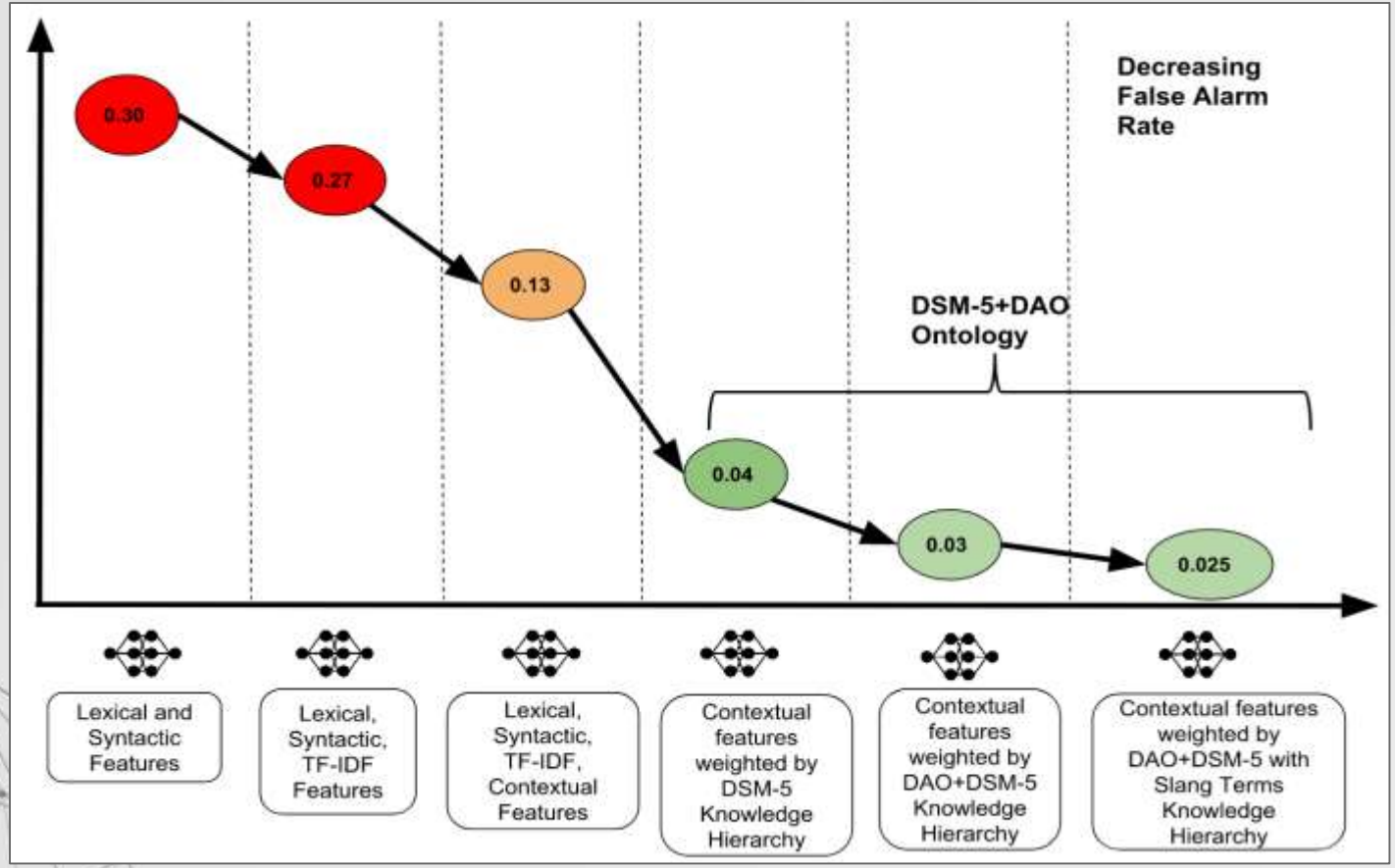
Continued: Explaining the prediction of mental health disorders (CIKM 2018)

- Explanation through word features created through Semantic Encoding & Decoding Optimization.
  - Semantic encoding of personal data into knowledge space
  - Semantic decoding of knowledge into personal data space.





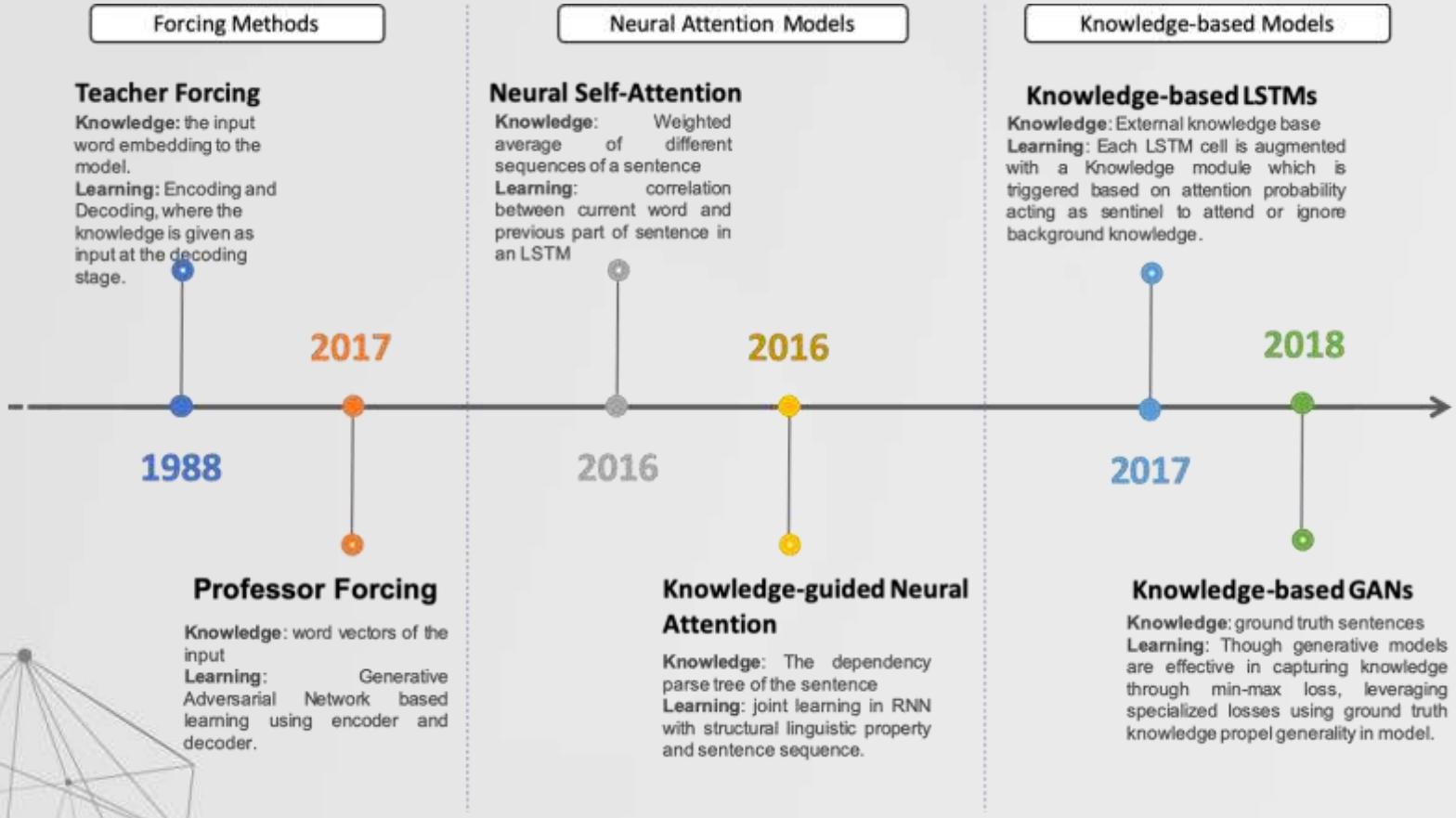
Continued: Explaining the prediction of mental health disorders (CIKM 2018)



Semi-Deep Infusion

In semi-deep infusion, external knowledge is involved through attention mechanism or learnable knowledge constraints acting as a sentinel to guide model learning.

SEMI-DEEP Infusion of Knowledge for Machine/ Deep Learning in Brief



External Knowledge through Attention

- A neural attention mechanism equips a neural network with the ability to focus on a subset of its inputs (or features):

    - **Hard Attention** or Position specific attention : location of important entities and relationship in the text are hard-coded in the model. Thus allowing efficiency in feature engineering, however, the model suffer from exposure bias.
    - **Soft Attention:** The model learns to *attend* to specific parts of the text while generating the word describing that part (following distributional semantics).
    - **Attention with Knowledge base:** background knowledge is integrated using an attention mechanism, which decide whether to attend to background knowledge and which information from KBs is useful.

External Knowledge through Learnable Constraints

- Learnable constraints are empirical thresholds (probabilistic value) learnt by the model which allows it to adaptively learn.

- It can be done in following ways:
  - Learning based on pre-structured axiomatic rules - axiomatic knowledge
  - Learning based on difference in content similarity - KL Divergence, Cross-entropy loss
  - Learning based on commonsense knowledge - ConceptNet
  - Learning over different permutations of text generated through synonyms, antonyms, and homonyms.

Template: fill in the blanks

_____ meant to _____ not to _____

Generative Model

It was meant to dazzle not to make it

Target:

It was meant to dazzle not to make sense

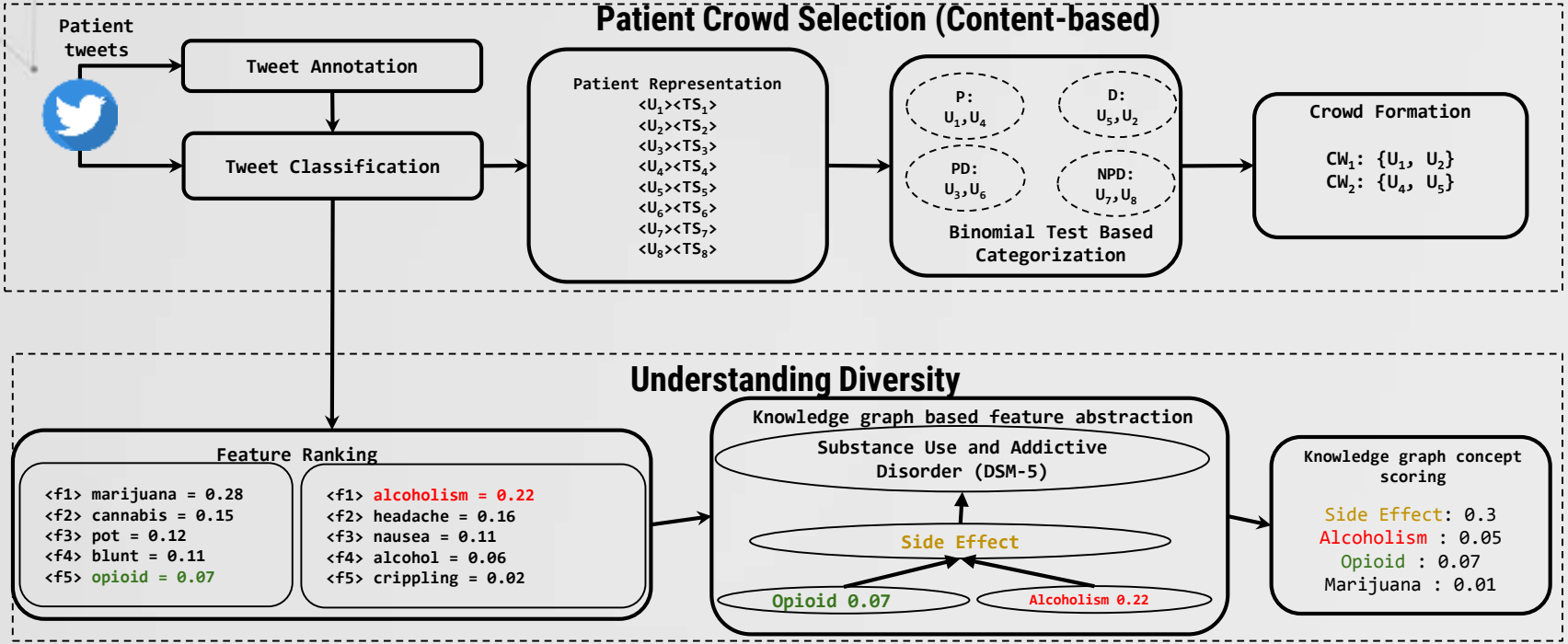
Infilling Content Matching through averaged KL Divergence

Learnable knowledge constraint module

Relevant research:

Explaining Mental Health Classification

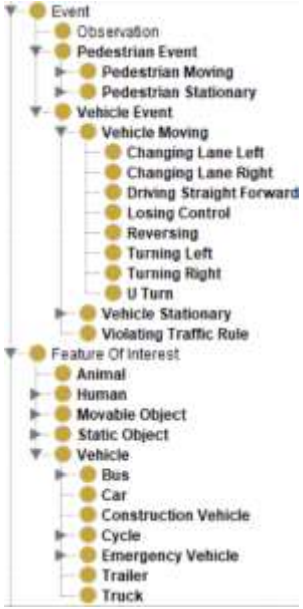
by adapting the prediction for wisdom of crowd

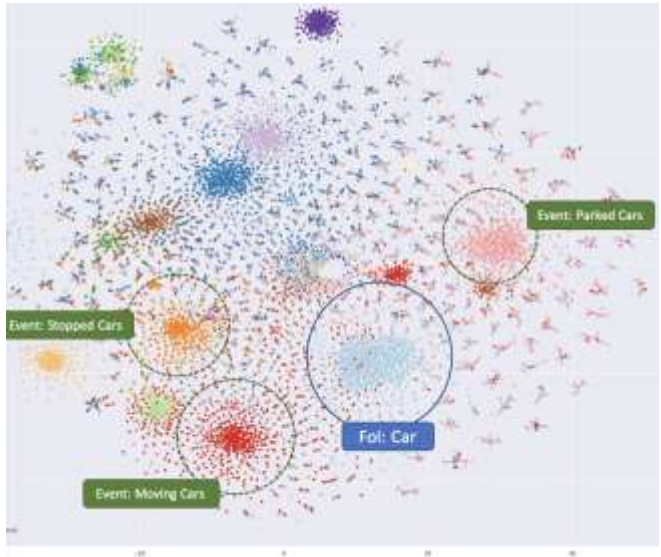


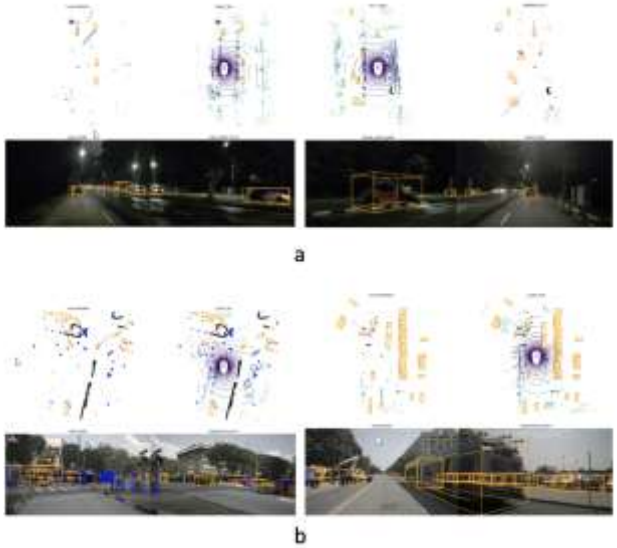
Bhatt, S., Gaur, M., Bullemer, B., Shalin, V., Sheth, A., & Minnery, B. (2018, December). Enhancing crowd wisdom using explainable diversity inferred from social media. In 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI) (pp. 293-300). IEEE.

Relevant Research:

Knowledge Graph Embeddings for Autonomous Driving







Scene KG

KG Embeddings of objects/events

Computed Scene Similarity

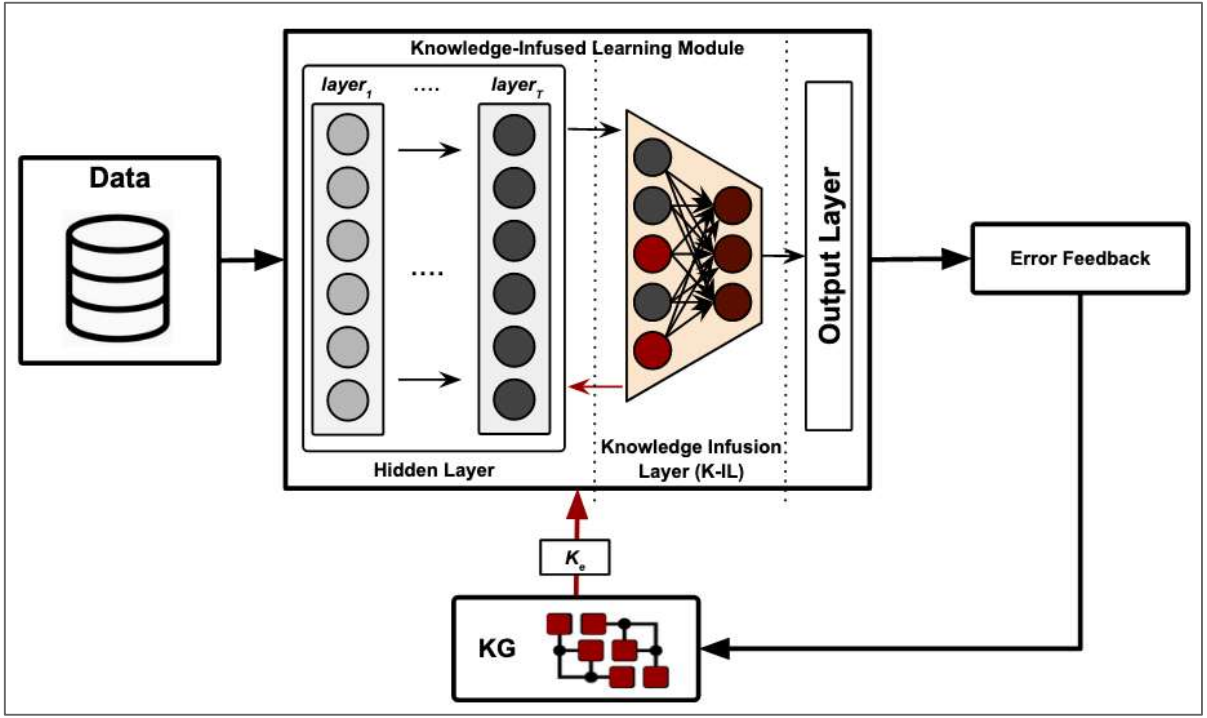
Wickramarachchi, Ruwan., Henson, Cory., and Sheth, Amit. An evaluation of knowledge graph embeddings for autonomous driving data: Experience and practice. In AAAI 2020 Spring Symposium on Combining Machine Learning and Knowledge Engineering in Practice (AAAI-MAKE 2020).

Deep Infusion

In deep infusion, external knowledge is incorporated within the neural network. The latent representation is transmitted between layers including hidden layers, while achieving enhanced abstractions.

DEEP Infusion of Knowledge for Machine/ Deep Learning

- Infusion can take place (i) **before the output layer** (e.g., Soft-Max), (ii) between hidden layers (e.g., reinforcing the gates of an NLM layer, modulating the hidden states of NLM layers)
- Classification **error** and **KG** determine the need for infusing knowledge. The Knowledge Infusion Layer incorporates the knowledge in the latent representation before output layer.



Ugur Kursuncu, Manas Gaur, and Amit Sheth. "Knowledge Infused Learning (K-IL): Towards Deep Incorporation of Knowledge in Deep Learning." AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering. Palo Alto, California, USA. 2020.

Open Research Questions:

Knowledge-Infused Learning

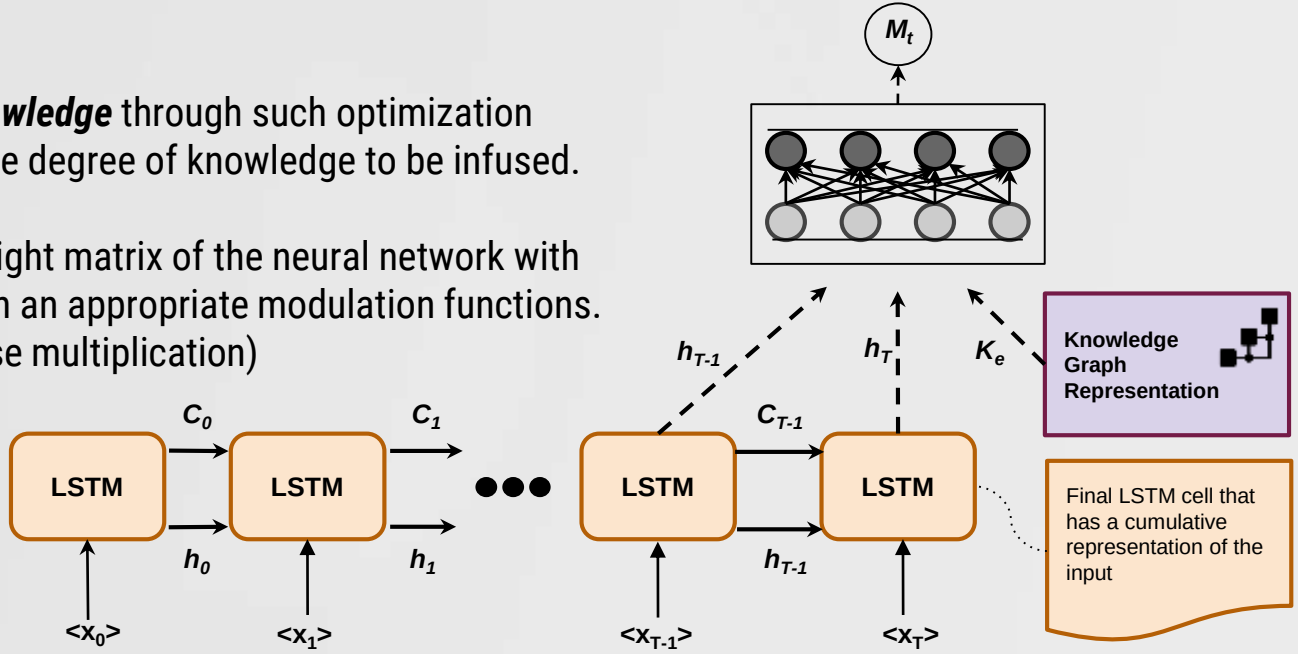
- How to decide whether to infuse knowledge or not at a particular stage in learning between layers, and how to measure the incorporation of knowledge?
- How to merge latent representations with knowledge representations, and how to propagate the knowledge through the learned representation?

The Learning is aligned with stratified knowledge, and knowledge structure and abstractions are significantly retained. Specific functions and how they can be operationalized.

- Knowledge-Aware Loss Function
- Knowledge Modulation Function

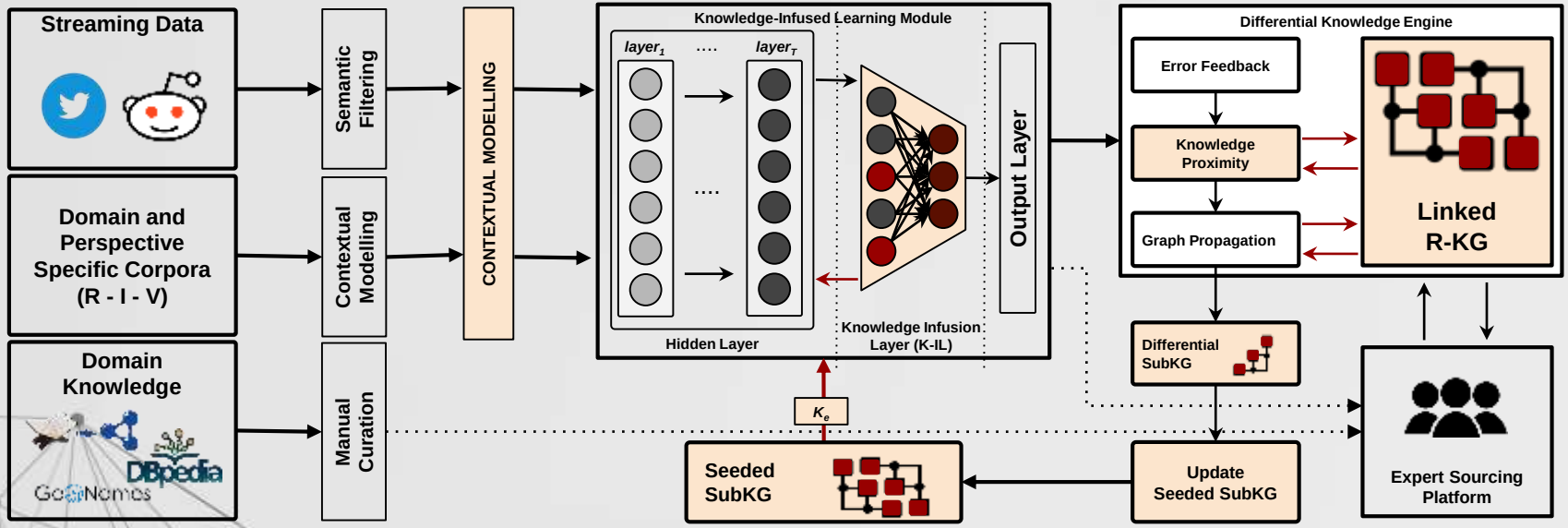
Knowledge-Infused Learning Layer

- Optimize the loss function as the difference between the latent representations(h_{T-1} h_T) of the last cell is reduced with respect to the knowledge representation, e.g., through KL divergence.
- Compute ***differential knowledge*** through such optimization approach; determining the degree of knowledge to be infused.
- Modulate the learned weight matrix of the neural network with the hidden vector through an appropriate modulation functions. (e.g., Hadamard pointwise multiplication)



Broader Perspective:

Knowledge-Infused Learning



FUTURE

Computing for Human Experience

*My vision for semantic technologies is for it to find a place in future more powerful AI resulting from a synergy between the **top-down processing** (symbolic AI) and **bottom-up processing** (statistical AI) in human brains.*

Knowledge Graphs

will play an increasing role in developing hybrid neuro-symbolic systems (that is bottom-up deep learning with top-down symbolic computing) as well as in building explainable AI systems for which KGs will provide scaffolding for punctuating neural computing.

KG: The Glue in Developing Hybrid AI Systems

SYMBOLIC AI
FORMAL

"Unreasonable effectiveness of small data"
in human decision making - can this be emulated to power top down processing?

STATISTICAL AI
CONNECTIONIST

"Unreasonable effectiveness of big data"
in machine processing &
powering bottom up processing

Explainability, Traceability, and Interpretability in AI

A hybrid approach integrating top-down with bottom up

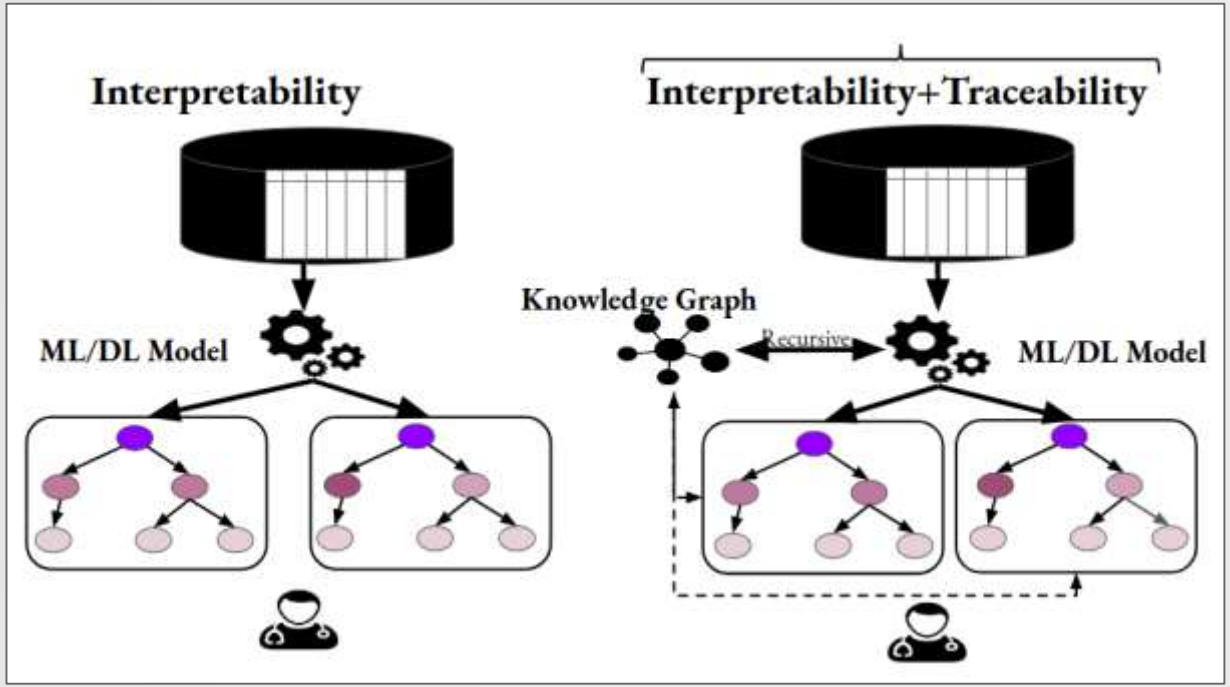


Figure: KG-infused Machine/ Deep Learning

PROMISING KG IMPACTS

ROBOTICS	COGNITIVE SCIENCE	SELF-DRIVING CARS	PERSONAL ASSISTANT
Cross-domain Knowledge 1) Observational (sensory data) and common-sense knowledge to perceive the surrounding environment 2) Knowledge representation to model the knowledge concerning the surrounding environment 3) Appropriate cross-domain knowledge reasoning mechanisms	**Human Intelligence** “Inject” human intelligence into AI assistants such as Amazon Alexa, utilization of cross-domain knowledge of social interactions, emotions and linguistic variations of natural language.	**Empathy and Morality** AI agents to mimic human emotions and decisions, we need to model human emotional knowledge of empathy, moral, and ethics.	**Personalization** Smart ***health*** agents are adapting to answer real-world personalized complex health queries in simple interactive language. Requires patients’ environmental knowledge, health data, and coordination with their healthcare physicians.

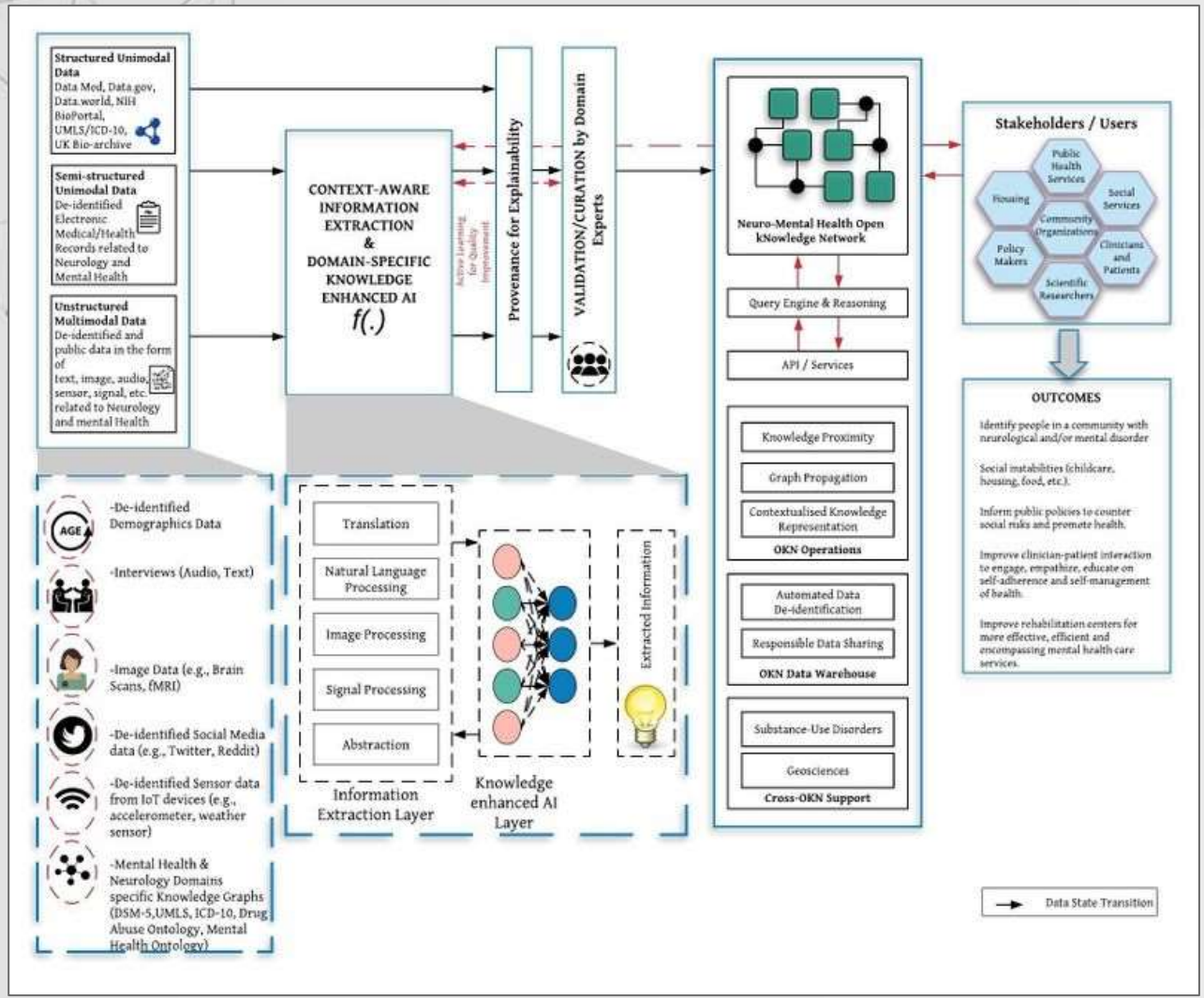


Figure: Illustrative example of OKN for neuro-mental domain.

CHALLENGES For KG

01 CAPTURING CONTEXT

02 DOMAIN-SPECIFIC KNOWLEDGE EXTRACTION

03 KNOWLEDGE ALIGNMENT

04 REAL-TIME KG FOR FAST DATA

05 QUALITY & VALIDITY OF KGs

06 ADAPTIVE KNOWLEDGE NETWORK

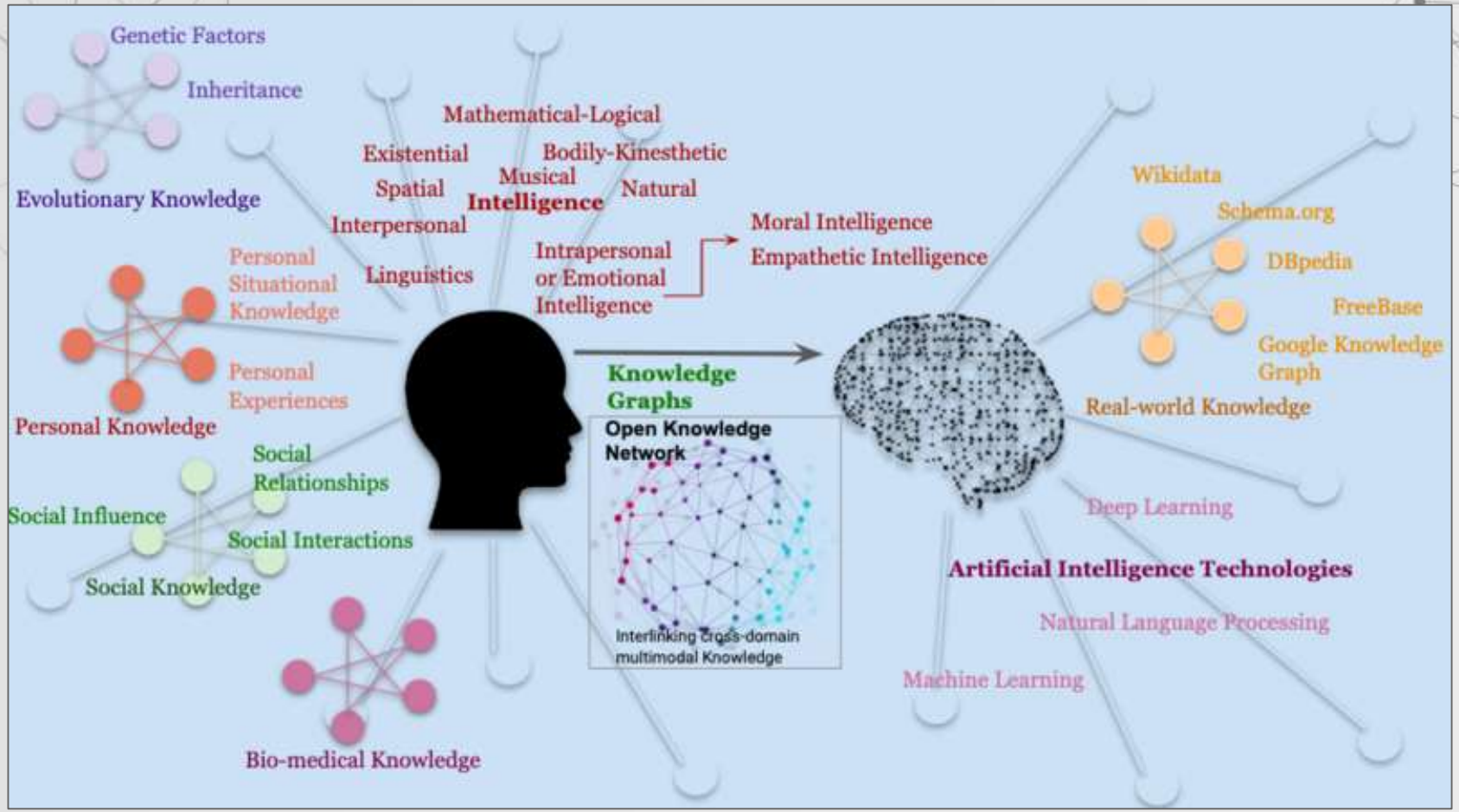


Figure: An expanding role of Knowledge in future AI systems.

Conclusion & Take away

“Data alone is not enough”:

https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf

Consider **combining data-centric**/bottom up/statistical learning **with knowledge-based**/top down techniques

- To improve understanding of simpler content
- To understand complex content and concepts
- To understand heterogeneous/multimodal content
- and a lot more

ai.sc.edu

References

- Valiant, Leslie G. "Robust logics." *Artificial Intelligence* 117.2 (2000): 231-253.
- Valiant, Leslie. *Probably Approximately Correct: NatureÕs Algorithms for Learning and Prospering in a Complex World*. Basic Books (AZ), 2013.
- Valiant, Leslie G. "A neuroidal architecture for cognitive computation." *Journal of the ACM (JACM)* 47.5 (2000): 854-882.
- Davis, Ernest, and Gary Marcus. "Commonsense reasoning and commonsense knowledge in artificial intelligence." *Commun. ACM* 58.9 (2015): 92-103.
- Xu, Keyulu, et al. "What Can Neural Networks Reason About?." *arXiv preprint arXiv:1905.13211* (2019).
- Marcus, Gary. "Deep learning: A critical appraisal." *arXiv preprint arXiv:1801.00631* (2018).