Enterprise-scale knowledge graphs
Panel: Enterprise-scale knowledge graphs

- Yuqing Gao, Microsoft
- Anant Narayanan, Facebook
- Alan Patterson, eBay
- Jamie Taylor, Google
- Anshu Jain, IBM
Knowledge Systems

Yuqing Gao
AI and Research
Microsoft
Knowledge systems: World graph

Search queries, views, click throughs, ...

World graph
- People
- Places
- Organizations
- Things
- Actions
...

Web pages, Web documents, Images, ...
Knowledge systems: Academic graph

Knowledge acquisition, search, recommendation ...

Authors, institutions, articles, conferences ...

Academic graph
- People
- Publications
- Fields of Study
- Venues
...
Knowledge systems: Work graph

Messages read/sent, Document author/shared, ...

Work graph
- People
- Groups
- Messages
- Activities
...

Emails, Messages, Documents, Meetings, ...
Infusing knowledge: From search to conversation

Search

Q&A

Enrichment

Relatedness

Conversation

Recommendations
Bing – knowledge in answers

area size of switzerland

Switzerland · Area
15,940 sq miles (41,285 km²)
About twice the size of New Jersey
I want to travel to NY 2 days before Thanksgiving, staying for a week.

Okay, booking a flight to JFK from November 20 to November 27. Where will you be flying from?

From San Francisco, and non-stop in first class.

Got it, I've found some flights for you ...

How about leaving in the afternoon?
How do we bring knowledge systems to life?

- Raw Data
  - Structured + Unstructured
- Knowledge creation
  - High-quality Semantically Organized
Challenges of scaled KGs

Building a small KG is easy - building a vast system like Satori is a huge challenge.

Coverage
Have we got the information we need?

Freshness
Is information up to date?

Correctness
Is our information accurate?

Three forces in constant conflict:

- Increased freshness and coverage
- Harder to ensure correctness

- Increased correctness
- Harder to ensure freshness and coverage

Correctness is always hard – what is true and correct? Particularly critical in today’s world.

Will Smith: Single entity, 108K facts assembled from 41 web sites. There are 200 Will Smiths on Wikipedia alone.
Active research and product efforts in knowledge

- Unsupervised knowledge extraction from unstructured data in open domain
- Knowledge graph semantic embedding
- Autonomous knowledge inference & verification
- Real-time knowledge graph with archiving
- Large scale entity linking and disambiguation
- Ultra-scale knowledge representations
- Knowledge system for multi-lingual
- Knowledge Precision vs Comprehensiveness
- ...
Thank you!
Why a Knowledge Graph for FB?

Facebook is known for the world’s largest social graph spanning billions of entities and trillions of assertions.

We’ve built technology over the last decade to enable rich connections between users and we’re eager to apply it to build a deeper understanding of not just people, but also the things that people care about.
What do we put in it?

The most **socially relevant** entities, concepts and words.

People, Celebrities
Places, Points of Interest
Movies, TV, Music
Sports
How large is it?

Deliberately focused and small to start with, US only.

~500M assertions
~50M entities
~100s of types
Knowledge of the world
People, Places, Things

Heidis Waffle House
Café & Restaurant

San Diego City
North Park Neighborhood
30th St Street
DoorDash Delivery
Zagat Review
TripAdvisor Review
Facebook Content

Attribute: adventurous, casual, sustainable, trendy
Dish: coffee and tea, bread, drink, parfait, cake, belgian waffle, gingerbread, liege waffle, turkey sandwich, fresh-squeezed lemonade, bacon waffle
Features: Credit cards, Takeout, Wifi, -Parking, -ADA
Meals: Breakfast, Lunch
Suggestions: liege waffle, lemonade
Telephone: (555) 987-1234
Hours: { ... }
Website: http://www.heidiswafflehouse.com
Use it to provide utility...

Recommend smart replies

Entity Detection

Easy sharing

Memory
... and delight the user!

Hey are you free now?
Yeah. Where are you?

You ready to go?
Yep, let's get a ride

Wanna catch up soon?
Let's meet up tmrw at 4pm

Can't wait to see you!
Yay!!
eBay Product Knowledge Graph

ISWC 2018

Alan Patterson, Structured Data
alpatterson@ebay.com
Nest 3rd Generation Learning Stainless Steel Programmable Thermostat

Our pick

$209.99

Free Shipping

List price: $249.00

Get it by Tuesday, Mar 8 from New York, United States

- New condition
- 30 days money back
- See full description

Sold by: inbeccommerce ($774)

58.1% Positive feedback

Quantity: 1

Buy It Now

Add to cart
Pipeline

**Graph** - Products, properties, value types, fitment, relationships, standards, variations, people, places, brands, companies, events, dates.

**Statistical Data Mining** - Discovery (over billions of listings)

**Buyer & Seller Data** - Queries, clicks & conversions, seller descriptions

**Knowledge Matching and Semantic Verification**

**Graph stats** - Expect around 100M products, >1Bn triples
Problems

Biggest Problem: Identity

Are these listings the same product

Identity depends on who’s asking

Buyer and Seller have different requirements
Knowledge Graph

1 Billion Things

70 Billion Facts
What is it used for

Introducing the Knowledge Graph: things, not strings
May 16, 2012
Ocean sunfish

Fish

The ocean sunfish or common mola is the heaviest known bony fish in the world. Adults typically weigh between 247 and 1,000 kg. The species is native to tropical and temperate waters around the globe. It resembles a fish head with a tail, and its main body is flattened laterally. Wikipedia

Mass: 2,200 lbs (Adult)
Scientific name: Mola mola
Length: 5.9 ft. (Adult)
Speed: 2 mph (Maximum, Adult)
Family: Molidae
Did you know: By basking on its side at the surface, the sunfish also allows seabirds to feed on parasites from its skin. eol.org

People also search for

Blobfish Sunfishes Sunfish Bluegill Molas

Sunfish

Sailboat

The Sunfish sailboat is a personal size, beach launched sailing dinghy utilizing a pontoon type hull carrying a lateen sail mounted to an unstayed mast. Sunfish was developed by Alcort, Inc. and first appeared around 1952 as the "next generation" improvement on their original boat, the Sailfish. Wikipedia

Hull weight: 120 pounds (54 kg)
Rig: Lateen
Total sail area: 75 square feet (7.0 m²)
Keel: Daggerboard
Draft: 2 feet 11 inches (0.89 m)
Year: 1953
Construction: Fiberglass
**Lateen**

A lateen or latin-rig is a triangular sail set on a long yard mounted at an angle on the mast, and running in a fore-and-aft direction. Dating back to Roman navigation, the lateen became the favorite sail of the Age of Discovery, mainly because it allows a boat to tack "against the wind."

[ Wikipedia ]

---

**Daggerboard**

A daggerboard is a retractable centreboard used by various sailing craft. While other types of centreboard may pivot to retract, a daggerboard slides in a casing. The shape of the daggerboard converts the forward motion into a windward lift, countering the leeward push of the sail.

[ Wikipedia ]
2018 Mazda MX-5 Miata
Sports car

Monterey
City in California

Euro
Currency

John Steinbeck
American author

Cannery Row
Novel by John Steinbeck
Book preview
16/106 pages available

Metaweb Technologies, Inc.
Company

Google
Challenges
Challenges

What is personally challenging

Long evolution
Bigger than my brain
So many stakeholders
Challenges

What is personally challenging
- Long evolution
- Bigger than my brain
- So many stakeholders

Wonderful Problems!
Challenges

What is personally challenging
Long evolution
Bigger than my brain
So many stakeholders

Survival Mechanism:
A Principled Approach
Managing Identity

Santos FC Before Pelé

Sailing

Cuba
Knowledge Graph @ Watson - Context

**Watson Jeopardy:** Fact based QA on Wikipedia and other sources

**Strategic IP Insight Platform - SIIP:** Interactive exploration & discovery on patents & journals

**Knowledge Integration Toolkit (KnIT):** Understanding all the worlds medical abstracts

**Watson Discovery Advisor:** OnPrem QA and interactive exploration based discovery
[youtu.be/pHfmh1Bi6aA](https://youtu.be/pHfmh1Bi6aA)

**Watson Discovery Service:** Cloud SAAS model for discovery on unstructured text (focused on developer)

**Several Watson Services:** Watson NLU, Retrieve and Rank, Document Conversion etc.
[ibm.com/watson/developercloud/services-catalog.html](https://www.ibm.com/watson/developercloud/services-catalog.html)
Discovery is about the non-obvious

Discovery creates new knowledge (hitherto unknown / unrecorded in domain docs, data sources). New knowledge is surprising and anomalous. It could be formally abstracted in several forms:

- **New link between entities**: A new side effect of a drug, a new potential emerging company as an acquisition target or sales lead, a non-obvious but relevant case / law applicable in a particular trial / proceeding, a person of interest in a terrorist attack. (Link Prediction, Relationship Discovery, Relationship Ranking)

- **A potential new important entity in the domain**: e.g. a new material for display technologies, a new investor for a particular investment area etc. (Entity Discovery, Entity Recommendation, Entity Ranking)

- **Changing significance of an existing entity**: it’s changing relationships/attributes/metrics. e.g. increasing stake by an investor in an organization, increasing interaction between a person of interest and some criminal in an intelligence gathering scenario, decreasing number complaints for a particular product or service in a retail / consumer scenario. (Trend analysis, Distribution Analysis, Anomaly Detection)

Search and Exploration look for knowledge already available in the knowledge sources available to the system. They are necessary for Discovery but not sufficient.

Confirmation of facts known to an SME of a domain is not discovery (listing all possible side effects of a drug which may have been mined from structured or unstructured data).
In a nutshell

• We haven’t focused on **ONE Global** knowledge graph
• We have a framework for anyone to build their domain specific KGs
• Offered as a part of Watson Discovery Advisor and now Watson Discovery Service, also consumed internally by services / solutions
• Clients In various sectors:
  – Banking and Finance
  – IT Services, Customer Service
  – Cyber Security
  – Scientific Discovery: Life Science, Oil & Gas, Chemicals & Petroleum
  – Defense
  – Space Exploration
  – Media and Entertainment
valuable things we learnt..

- polymorphic stores is the answer, stop fighting for one store
- build surprise into your model, into your api
- evidence is primitive to the system, which every node and edge maps to
- push entity resolution to runtime through context
- every new piece of knowledge should know its cascade effect

Detailed description can be found at talks.anshu.co
Open problems that haunt us

- modeling and analyzing changing knowledge
- merge relationships discovered from unstructured data with known relationships
- federation of global, domain specific and customer specific knowledge
- incremental update of global knowledge on horizontally scaled stores
- scanned PDF extraction ..garbage in garbage out
Thank you.