INFO216:
Knowledge Graphs

Andreas L. Opdahl
<Andreas.Opdahl@uib.no>

Session S11: Enterprise Knowledge Graphs

- Themes:
  - Open Knowledge Graphs *(S9-S10)*
    - ...or Linked Open Data resources / datasets
    - Wikidata, DBpedia, GDELT, EventKG GeoNames, WordNet, BabelNet...
  - Enterprise Knowledge Graphs (EKGs) (→ *S11)*
    - Google's knowledge graph
    - Amazon's product graphs
    - others (← F1)
    - News Hunter's infrastructure and architecture

Readings

- Sources:
  - Blumauer & Nagy (2020):
    Knowledge Graph Cookbook – Recipes that Work
    (parts 2 and 4)
- Resources in the portal, including:
  - lecture slides
  - “Things, not Strings” – the original Google blog post

Knowledge Graphs:
Is anyone really using this?

Is anyone really using this?

Yes!

- But...

Is anyone really using this?

Yes!

- But...
  - not quite as in the semantic web vision
  - not quite as in the LOD vision either
- Knowledge graphs are (additionally) becoming:
  - company internal
  - based on other technologies
    - such as general graph databases
  - not always linked to the LOD cloud

Is anyone really using this?

Yes!

- But...
  - not quite as in the semantic web vision
  - not quite as in the LOD vision either
- Knowledge graphs are (additionally) becoming:
  - company internal
  - based on other technologies
    - such as general graph databases
  - not always linked to the LOD cloud

Many of these ideas
are widely adopted too, such as:
- microdata / schema.org
- RDF / SPARQL / … for semantic data exchange
- graph representations in general

Is anyone really using this?

Yes!

- But...
  - not quite as in the semantic web vision
  - not quite as in the LOD vision either
- Knowledge graphs are (additionally) becoming:
  - company internal
  - based on other technologies
    - such as general graph databases
  - not always linked to the LOD cloud

Similar ideas, adapted to new uses and business contexts, using a combination of standard and other technologies


Tencent 腾讯
UniProt
USGS
Google
Bing
Alibaba.com
Baidu 百度
PubMed
facebook
ANTONI VAN LEEUWENHOEK FOUNDATION
MAASTRO
DEUTSCHE NATIONAL BIBLIOTHEK
europeana
BBC
The New York Times
GETTY
NXP
REUTERS
KUNGL. BIBLIOTEKET
National Library of Sweden
RENAULT
EPA United States Environmental Protection Agency
IOS Press
European Commission
Walmart
SIEMENS
BEST BUY
Deloitte.
SPRINGER NATURE
ELSEVIER
amazon.com
accenture

Google's Knowledge Graph

Google's Knowledge Graph

- Google Knowledge Graph (from 2012)
  - "Things, not Strings"
  - seeded from Freebase
  - facts from Wikipedia, Wikidata, CIA World Factbook
    - a growing number of other sources
  - enriched by natural-language parsing (NLP)
    - Google's Knowledge Vault
  - used internally for many purposes
  - visible in Google Search results (Knowledge Panels)
  - question answering in Google Assistant / Home

Caution: *The public documentation is limited, so this is compiled based on presentations, technical notes, forums etc.*

Google's Knowledge Graph

- Coverage:
  - claimed
    - 18 billion facts (18G, norsk: 18 milliarder) about 570 million entities *soon after start*
  - 70 billion facts claimed in (2016)
  - 500 billion facts about five billion entities (2020)
    - ...perhaps 3 times the size of the LOD cloud
  - from English to multiple languages
- Critiques:
  - source attribution, incl. Wikipedia / Wikidata

Caution: *The public documentation is limited, so this is compiled based on presentations, technical notes, forums etc.*

Google's Knowledge Vault Project

- Google Knowledge Vault
  - extends the Knowledge Graph
  - covers resources not from open semantic datasets
  - facts extracted from the whole web
    - NLP of text documents
    - HTML trees and tables
    - human annotated pages (e.g., schema.org)
  - probabilistic reasoning
    - graph-based priors
    - knowledge fusion

Caution: *The public documentation is limited, so this is compiled based on presentations, technical notes, forums etc.*

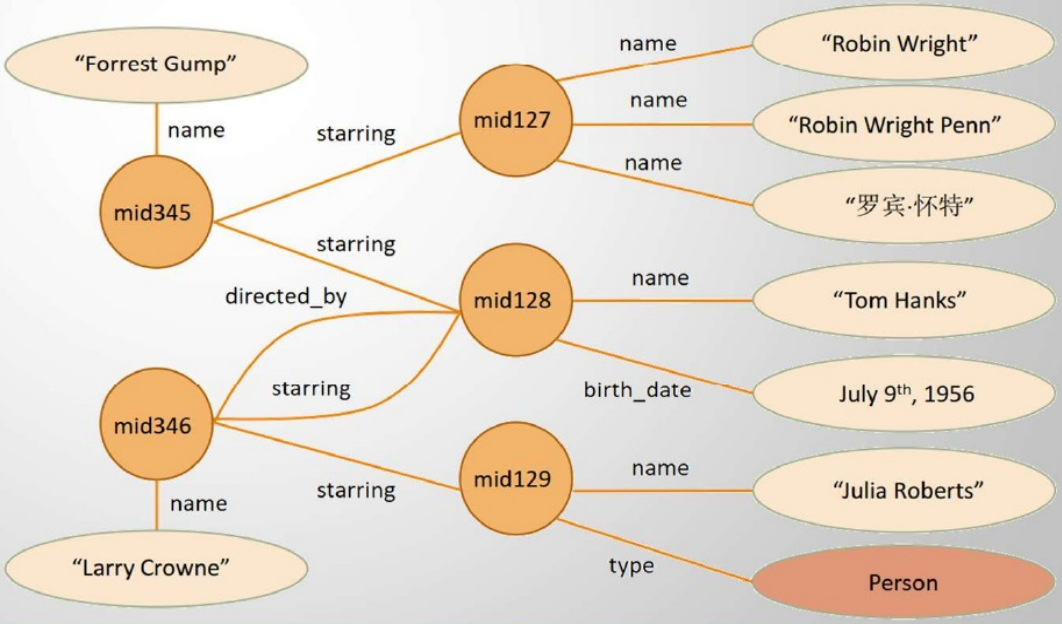
Amazon's Knowledge Graph

Amazon's ambition

- Let shoppers find the best products that fit their needs
  - allow greater variation in search terms
  - allow complex queries
- Structure all of the world's information as it relates to everything available on Amazon
- Describe every product on Amazon
  - concrete and abstract concepts
  - products and non-products
  - link different entities
- Enriched customer experience
  - visit Amazon to see what's new or interesting
  - discover ways to simplify and enrich their lives

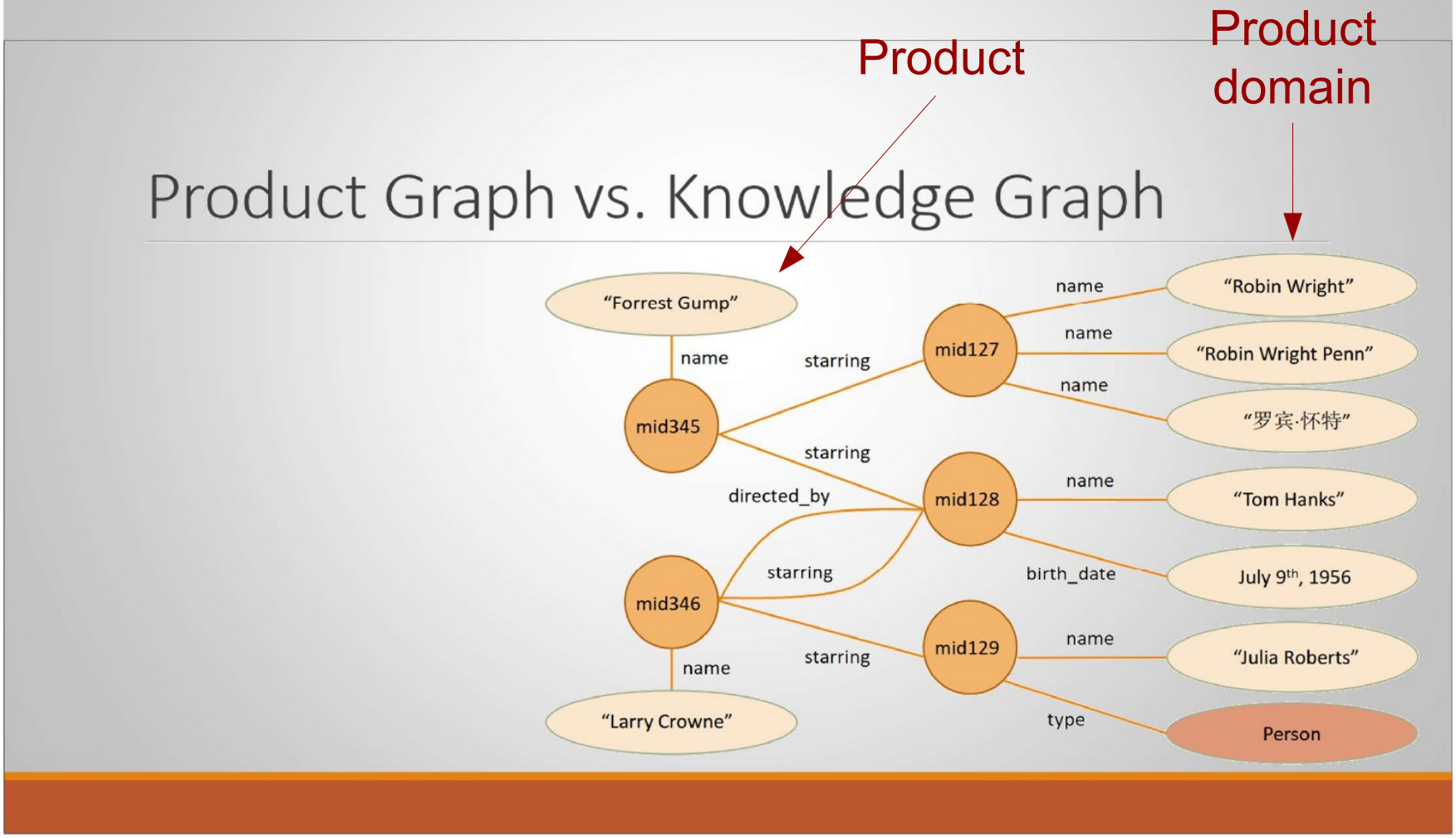
Amazon

Product Graph vs. Knowledge Graph



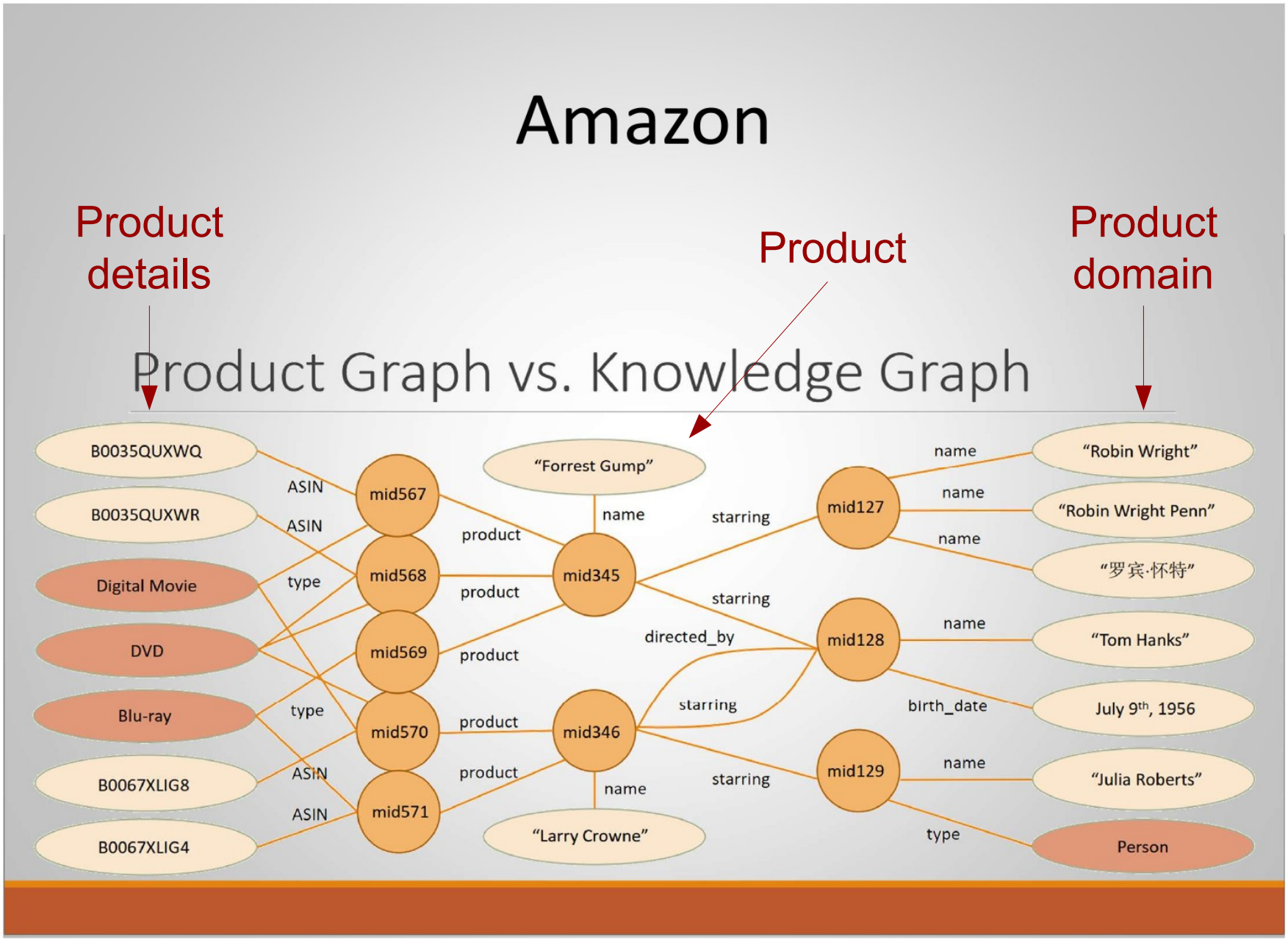
Frank van Harmelen (2018): Keynote at CAiSE'18

Amazon

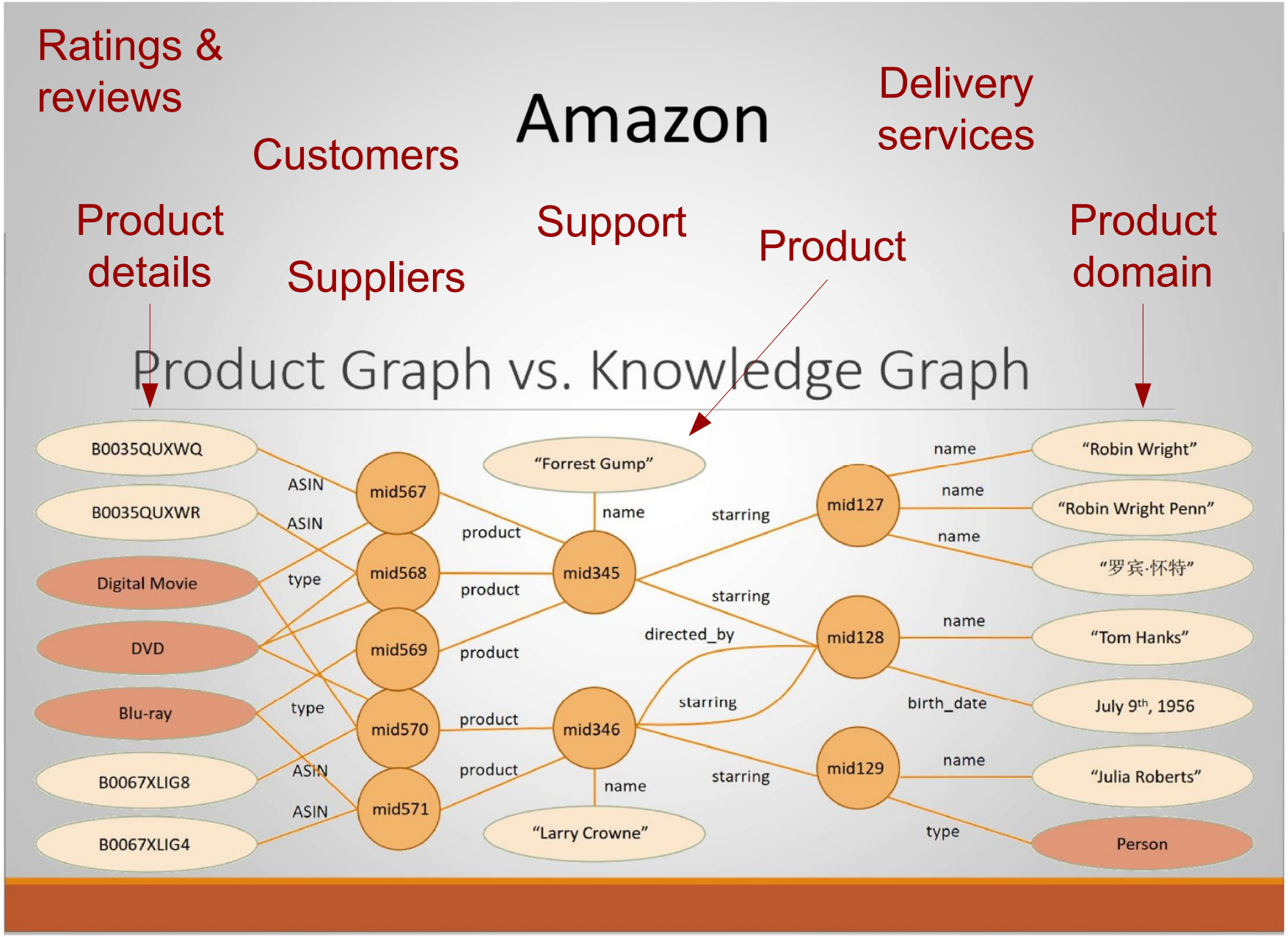


Frank van Harmelen (2018): Keynote at CAiSE'18

Amazon



Frank van Harmelen (2018): Keynote at CAiSE'18



Frank van Harmelen (2018): Keynote at CAiSE'18

Challenges

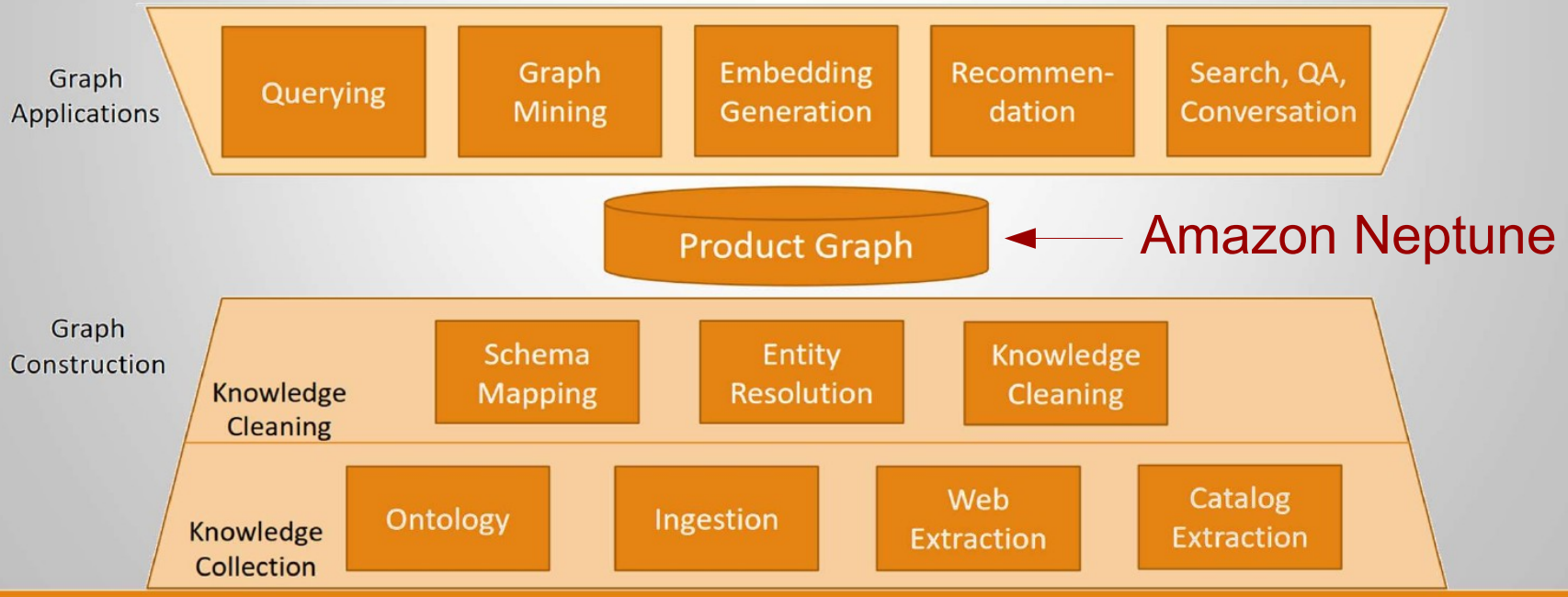
- Ingest product-related information from Amazon's detail pages and from the Internet at large
  - product information is largely unstructured
  - trustworthiness of sources
- Machine learning techniques for
  - knowledge extraction, linkage and cleaning
  - distantly supervised learning
    - train on more structured subset of data
    - run on larger unstructured data space
  - open information extraction
  - graph mining techniques to identify interesting hidden patterns (buying product-X buying product-Y)

Amazon

"We aim at building an authoritative knowledge graph for all products in the world"

Xin Luna Dong, Amazon, at WSDM conf, Feb 2018

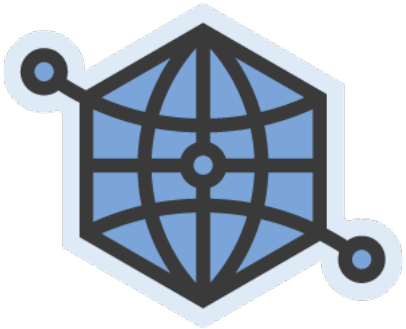
Architecture



Facebook's Social Graphs

Facebook's “Open” Graph Protocol (OGP)

- Including resources (in particular web pages), through their IRIs, in social graphs
  - targetting webmasters and content-management system (CMS) developers
- @prefix og: <http://ogp.me/ns#>
- Main properties:
  - required: og:title, og:type, og:image, og:url
  - optional: og:audio, og:description, og:determiner, og:locale, og:locale:alternate, og:site_name, og:video
    - ...some of them combines with more specific ones
  - *...markup with RDFa <meta>-tags*



OGP uses

- Uses:
  - originally developed by Facebook to extend the "Likes" mechanism to resources outside Facebook
  - also taken up by some other graph maintainers (claim: used by Google)
  - publishing side:
    - IMDb, Microsoft, Rotten Tomatoes, Yelp

Caution: *The public documentation is limited, so this is compiled based on presentations, technical notes, forums etc.*

OGP resource types

- <meta property="og:type" content=”ResType" />
- Some predefined resource types for:
  - music: music.song, music.album, music.playlist...
  - video: video.movie, video.episode, video.tv_show...
  - others: article, book, profile, website
- Each predefined resource type has further type-specific properties, e.g.,
  - music:duration, music:album:track, music:musician
- Data types:
  - boolean, date/time (ISO 8601), enum, float, integer, string, URL

Facebook's Graph API

- Letting external applications access the information in Facebook's social graph
  - inspired by *social networks*
- *Nodes* represent “things”: *User, Photo, Page, Comment*
- *Edges* represent connections between the "things":
  - Users' *friends*, Pages' *photos*, Photos' *comments*...
- *Fields* contain information about the "things":
  - the *birthday* of a User, the *name* of a Page...
- *Seriously restricted since version 2.0... (Privacy!)*
  - *the idea remains important*
  - *open, user-owned alternatives are emerging*
    - *GNU social (StatusNet), Diaspora...*

Facebook Graph API

- *REST*-based (REpresentational State Transfer)
  - an example of a *web service / web API*
  - all nodes have IRIs
  - GET, POST, DELETE over HTTP
- GET graph.facebook.com/facebook/picture?redirect=false
  - this is sent over HTTP (at least):
    GET /facebook/picture?redirect=false HTTP/1.1
    Host: graph.facebook.com
- Many API operations are based on *access tokens*
  - returned by *Facebook login*
  - mandatory for POST and DELETE
  - *friends' information must be explicitly granted*

Facebook Graph API

- Most HTTP-requests go to:
  - http://graph.facebook.com/...
  - http://graph-images.facebook.com/...
- Node paths:
  - GET graph.facebook.com/{node-id}
- Edge paths:
  - GET graph.facebook.com/{node-id}/{edge-name}
- With access token:
  - GET graph.facebook.com/me
- POST and DELETE are also used

Try it out: https://developers.facebook.com/tools/explorer

Wolftech's
News Hunter