Semantic Web and LLM-based Chat Bot Symbiosis

Kingsley Idehen,
CEO, OpenLink Software

OPENLINK SOFTWARE
Making Technology Work For You®

Presentation Goal

Help you understand why Hypertext is so important to computing (past, present, and future) via its practical impact on:

- Data Representation & Access
- Information Representation & Access
- Knowledge Representation & Access

Presentation Overview

- Understanding Data
- Hypertext Overview
- Semantic Web
  - Linked Data
  - Ontology
  - Knowledge Graphs
- Addressing User Interface Challenges with Large Language Models
- Large Language Model (LLM) based Bots & Semantic Web Symbiosis

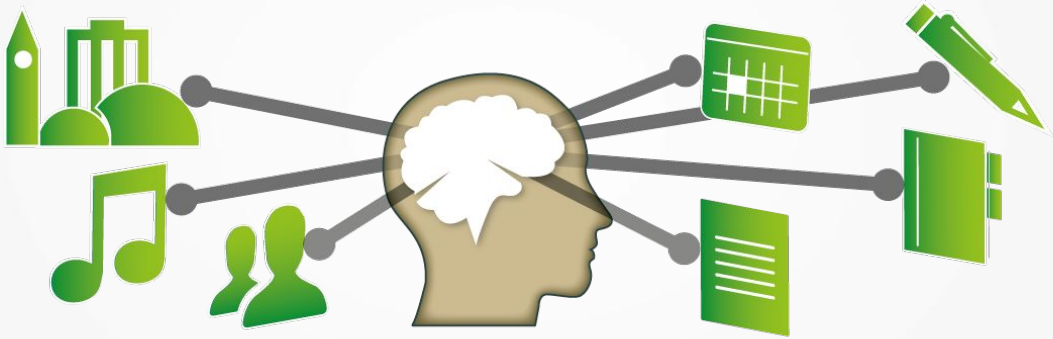
Understanding Data

What is Data?

Data is how we express Observation in reusable form.

What is Observation?

Observation is the Perception of Relationships between Entities.



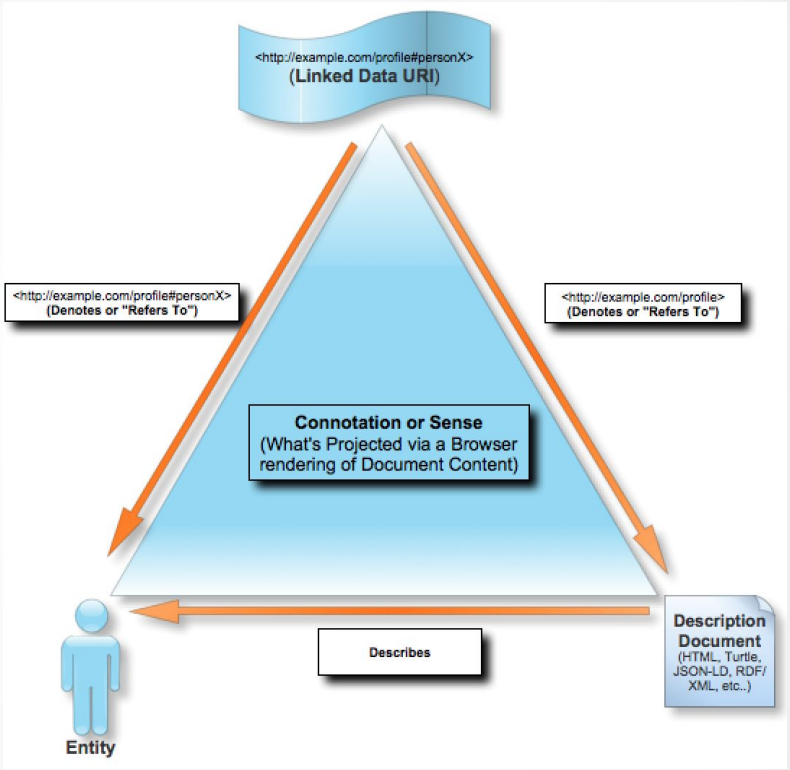
PEOPLE, PLACES, MUSIC, DOCUMENTS, CALENDARS, DIARIES, ADDRESS BOOKS & MORE...

What is an Entity?

An Entity is a Distinctly Identifiable Thing

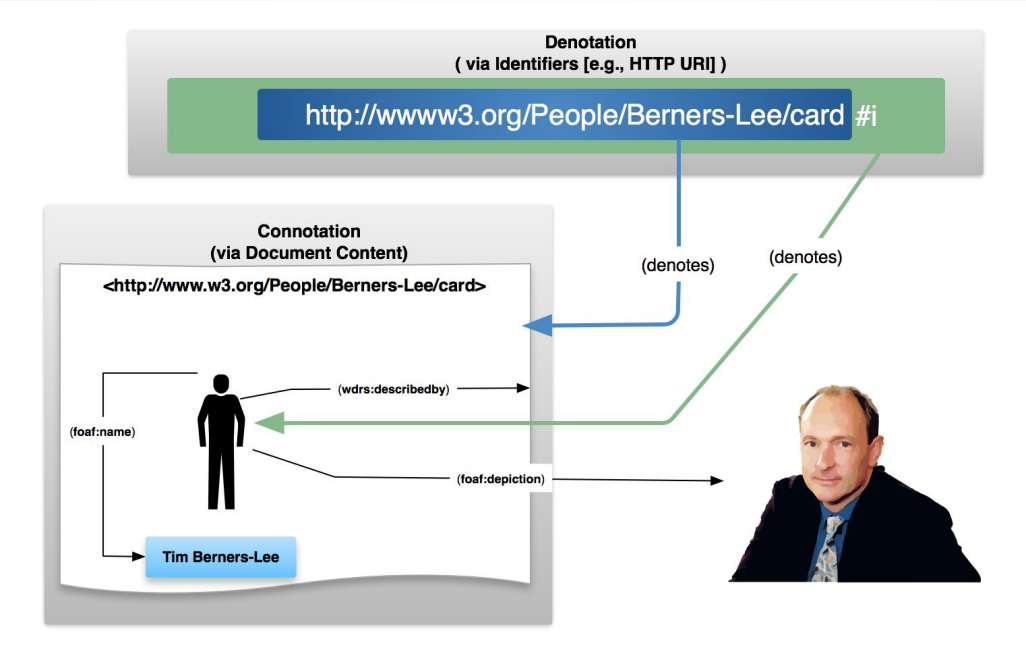
How is an Entity Identified (Named) ?

An Entity is Identified (or named) through the combined effects of Identifier based denotation (signification) and document content based connotation (description).



How Does Entity Identification Work?

Through interpretation that's driven by sign [denotation] -> description [connotation] based indirection.



How is an Entity Denoted?

An Entity is Denoted (Signified) through the use of an Identifier.

What is an Identifier?

An Identifier is a Sign (or Token) that Signifies (Denotes, or

“Stands For”) an Entity

Identifier Types?

Quoted Literals such as:

“Kingsley Idehen” or ‘Kingsley Idehen’

Absolute References:

<https://linkedin.com/in/kidehen#this>

Relative References:

<#KingsleyIdehen>

How is an Entity Described?

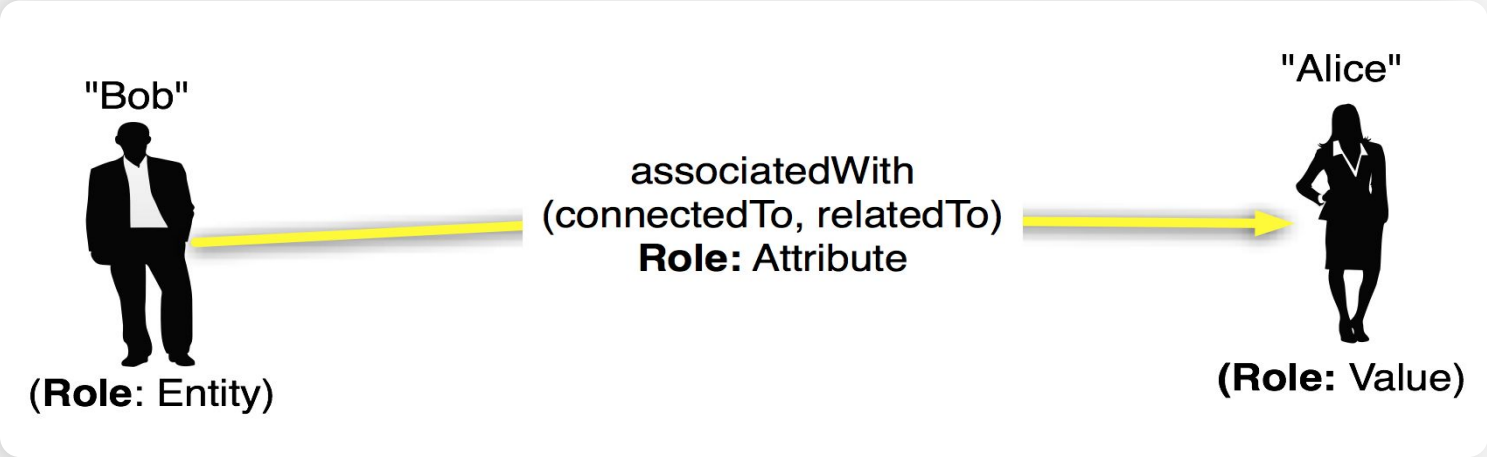
Through entity relationships that are represented in reusable form via document content (sentences and statements).

What is a Relationship?

A Relationship is an Association between two or more Entities, where each has a specific Role.

What is a Relationship Role?

A Relationship Role is a
Function performed by an Entity
in a Relationship

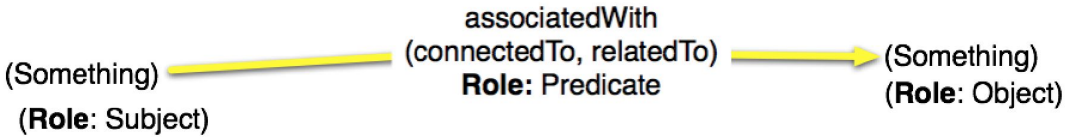


Relationship Role Types?

- **Entity Attribute Value EAV**
  - ✔ **Entity** -- observation focal point
  - ✔ **Attribute** -- observation attribute name (relationship type determinant)
  - ✔ **Value** -- observation attribute value
- **RDF (WC3's Resource Description Framework)**
  - ✔ **Subject** -- observation focal point
  - ✔ **Predicate** -- observation attribute name (relationship type determinant)
  - ✔ **Object** -- observation attribute value

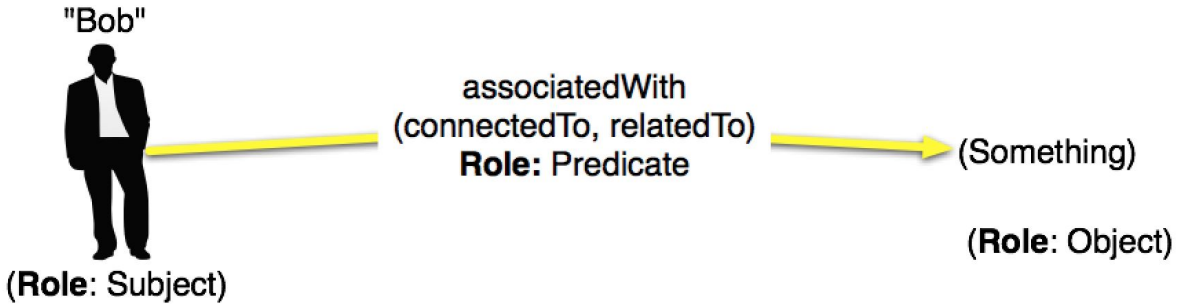
Relationship Role: Predicate

The Relationship Predicate is the Connector that associates an observation focal point (Subject) with something, in the form of an observation value (Object).



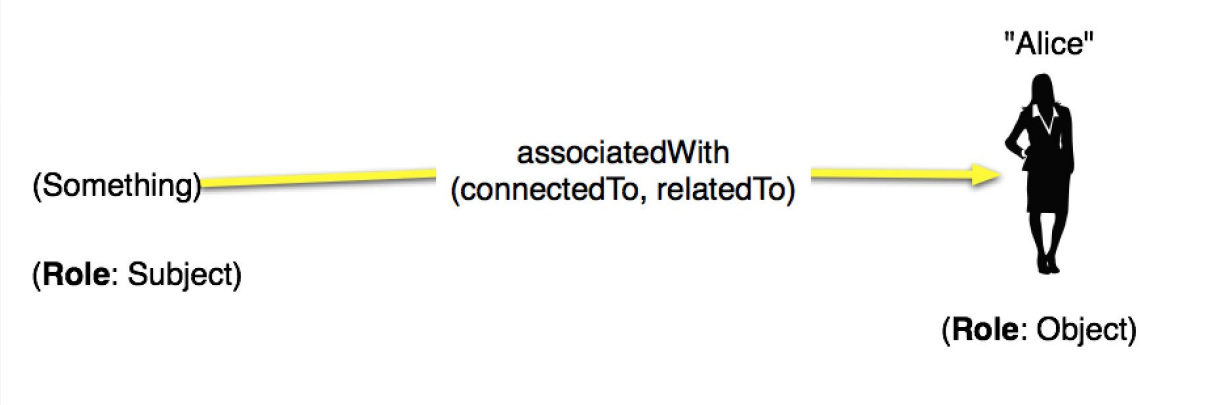
Relationship Role: Subject

Actual Entity being Observed



Relationship Role: Object

Value associated
with an observation focal point (Subject)
via a Relationship Predicate.



Types of Values?

- Untyped Literals (Strings)
- Typed Literals
  - ✔ Numbers
  - ✔ Dates
  - ✔ Booleans
  - ✔ Etc.
- References (Local and Global Hyperlinks)

How are Relationships Expressed?

Relationships are Expressed using a Language, i.e., a system of signs [for denotation], syntax [arrangement of signs to form sentences], and entity relation semantics [meaning of relationship roles] for encoding and decoding information.

Example:

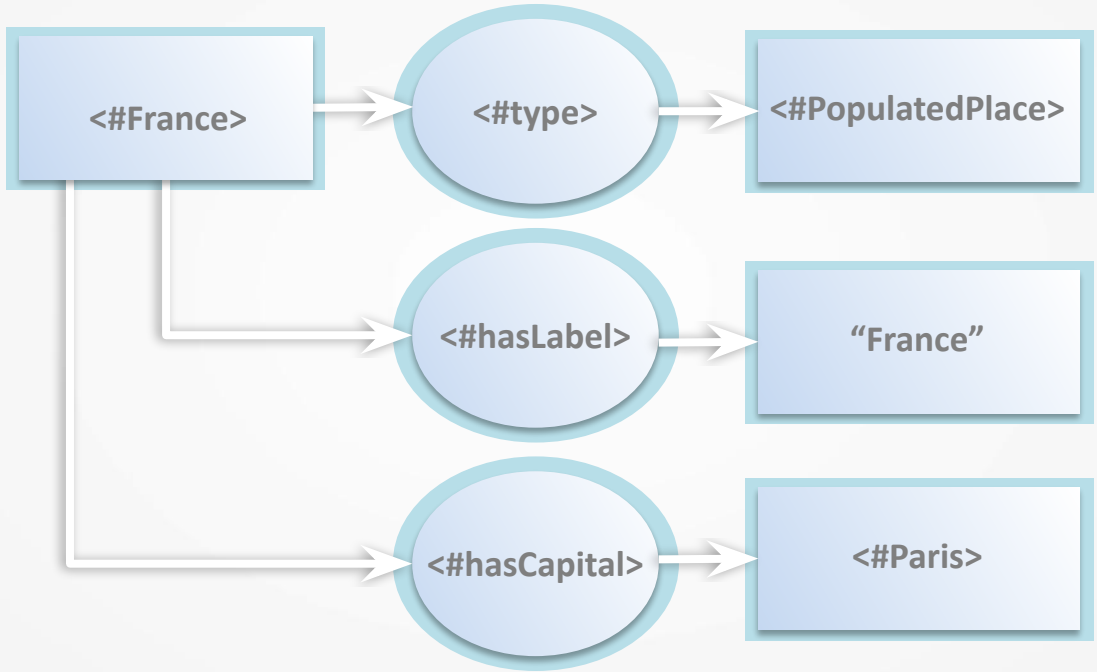
Subject, Predicate, Object –
Used by W3C's Resource Description Framework (RDF) and Natural Language.

How Are Entity Relationships Represented ?

Entity Relationships are Represented using notations associated with a specific language. Examples include:

- Graphic Notations used in Graph Diagrams
- Linear Notations e.g., RDF statements using JSON, JSON-LD, Turtle etc.
- Tables e.g., CSV files, Spreadsheets, and SQL Relational Database Management Systems).

Entity Relationship Diagram



Turtle Notation Based Entity Relationship Statements

<#France> <#Type> <#PopulatedPlace> .

<#France> <#hasLabel> ”France” .

<#France> <#hasCapital> <#Paris> .

<#Paris> <#Type> <#PopulatedPlace> .

<#Paris> <#hasLabel> ”Paris” .

<#PopulatedPlace> <#Type> <#Place> .

Entity Relationship Tables

Delimiter: e.g., Comma
Identifier Quote Character: Double-quotes
Relation Header Row: Entity,Attribute,Value
Relation Body

Example:
“Entity”, “Attribute” “Value”
“France”, “Type” “PopulatedPlace”
“France” , “hasLabel” “France”
“France” , “hasCapital” “Paris”

Statement Representation: Spreadsheet Tables

Entity (Subject)	Attribute (Predicate)	Value (Object)
#France	#Type	#PopulatedPlace
#France	#hasLabel	“France”
#France	#hasCapital	#Paris
#Paris	#Type	#PopulatedPlace
#Paris	#hasLabel	“Paris”
#PopulatedPlace	#Type	#Place

Hypertext

Hypertext Timeline

Year	Event	Key Players
1945	"As We May Think" essay proposes the idea of the Memex, a machine that can store and link information on microfilm	Vannevar Bush
1963	The term "hypertext" is coined as part of the Xanadu project, which aims to create a global network of linked documents	Ted Nelson
1967	The Hypertext Editing System (HES) is developed as one of the first systems that allows users to create and edit hypertext documents on a computer	Ted Nelson and Andries van Dam
1968	The NLS (oN-Line System) is demonstrated as a system that enables users to create and manipulate hypertext documents using a mouse, a keyboard, and a display	Douglas Engelbart
1989	The World Wide Web is created as a way to share information among researchers using a simple hypertext language (HTML), a protocol for transferring data (HTTP), and a system for identifying resources (URL)	Tim Berners-Lee
1990s	The Web becomes popular with the advent of graphical web browsers, such as Mosaic and Netscape Navigator	Various developers and users

Internet

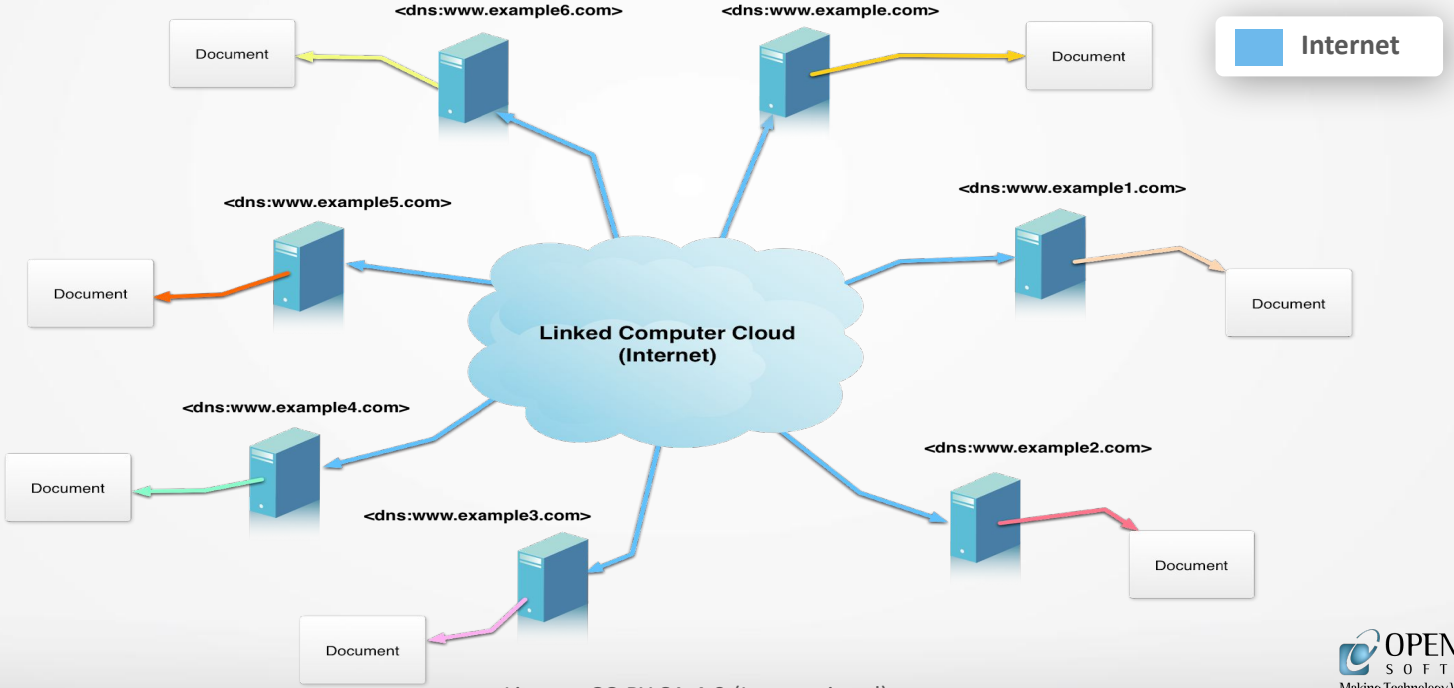
a/k/a

Linked Computer Network

DNS based Linked Computer Network (Internet)

Linked Computer Network (e.g., Internet)

1. Computer (DNS CNAMES) Names are Data Source Name
2. Actual Data Model and Data Access is Local and Machine OS hosted App. specific.



World Wide Web

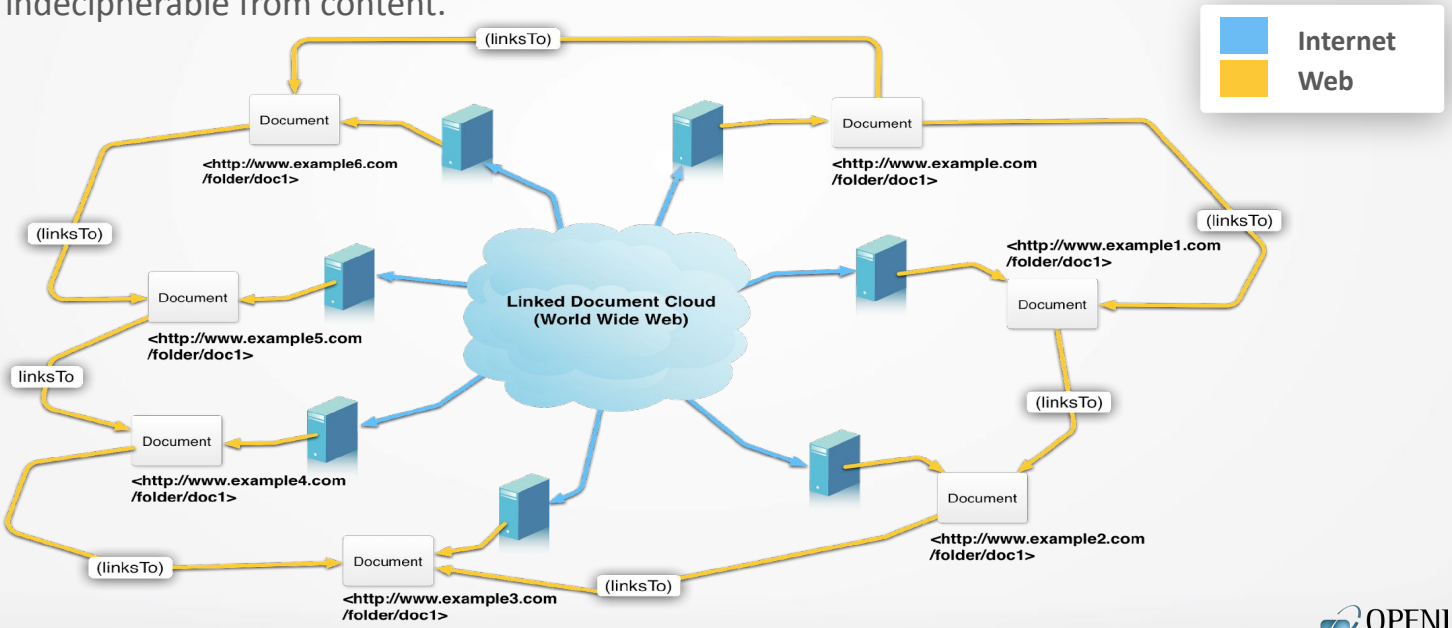
a/k/a

Linked Documents Networks

HTTP based Linked Document Network (Web 1.0 & 2.0)

Linked Document Network (e.g., World Wide Web)

1. Computer (DNS CNAMES) Names become irrelevant.
2. Document Locators / Addresses (HTTP URLs) are Data Source Names (DSNs).
3. One kind of Relation i.e., "LinksTo" is what connects the Documents.
4. To machines: actual Data Model, Entity Relation Semantics, and Representation Notations are indecipherable from content.

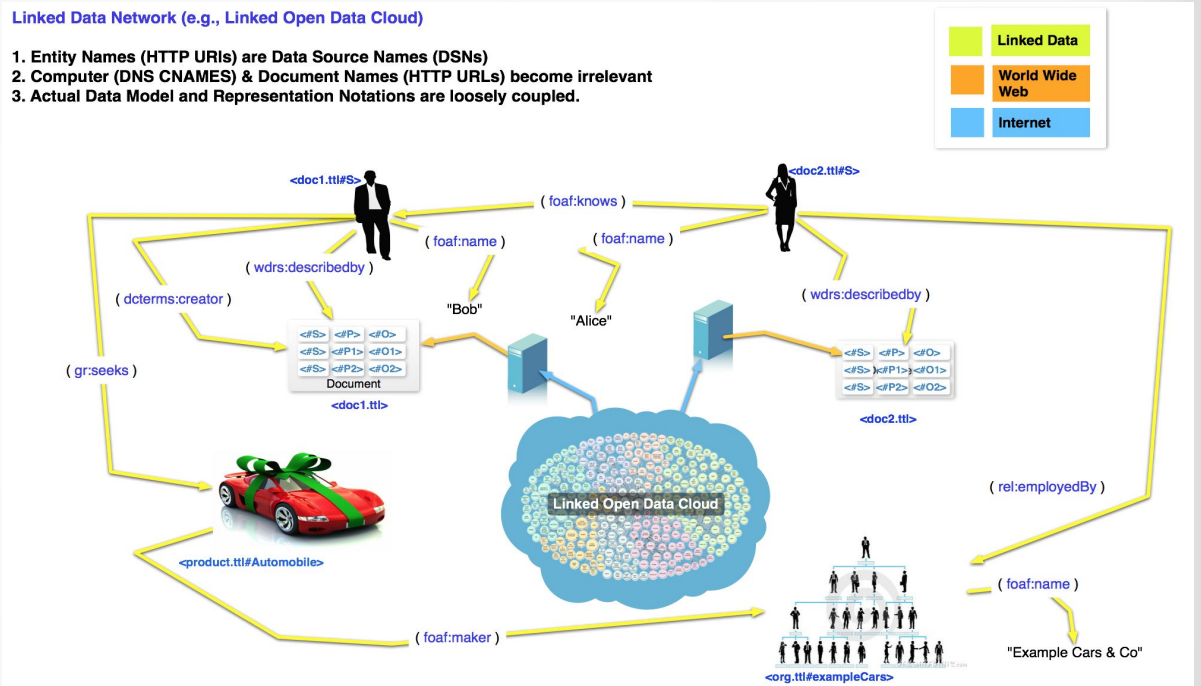


Linked Data

(WEBBY STRUCTURED DATA)

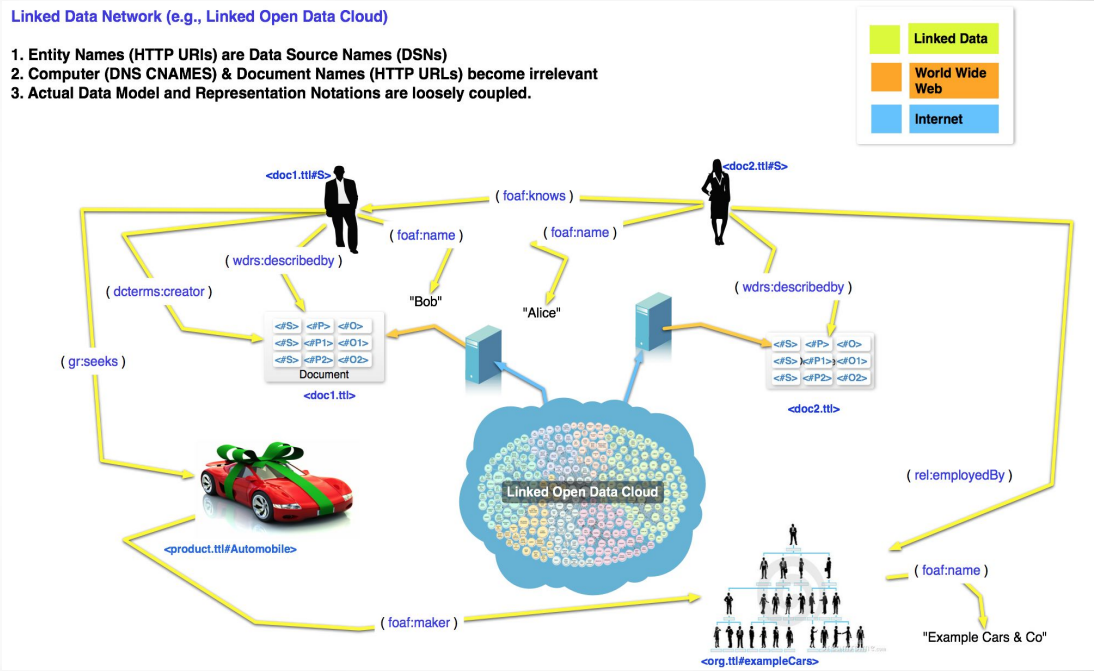
What is Linked Data?

A **principled** approach to **Structured Data Representation** that manifests as a **Web of Data.**



How Do You Create Linked Data?

- **Name** anything using a **Hyperlink**
- **Describe** everything using **Structured Sentences** where the **Subject, Predicate, and Object** (optionally) are named using a **Hyperlink.**

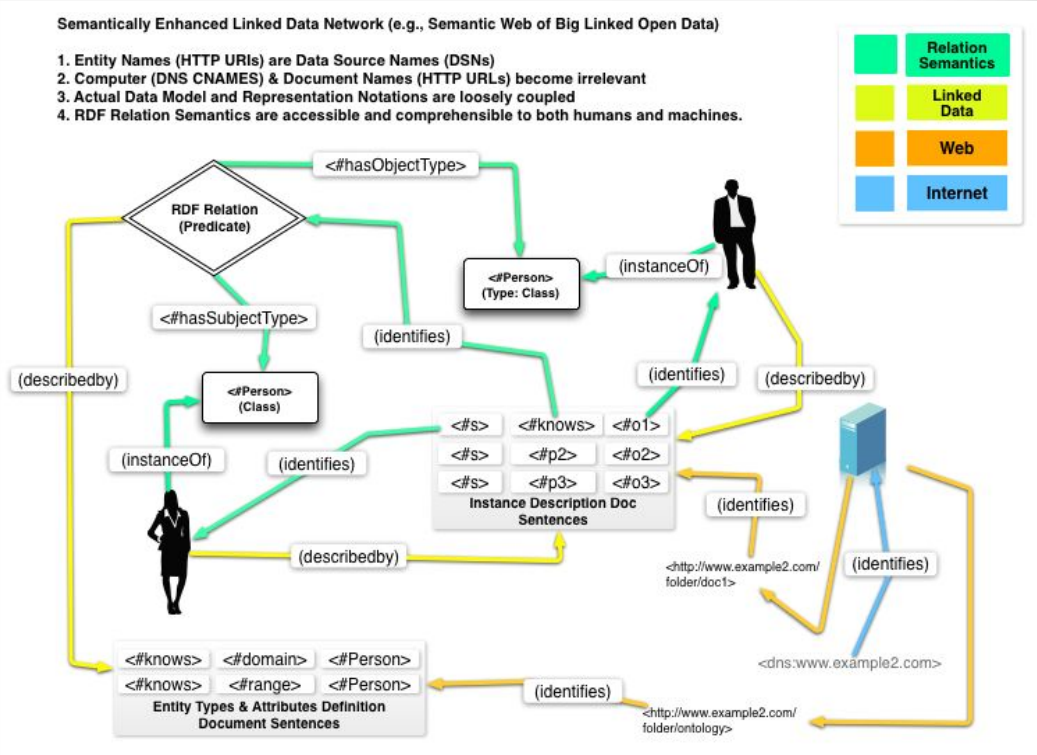


Ontology

What is an Ontology?

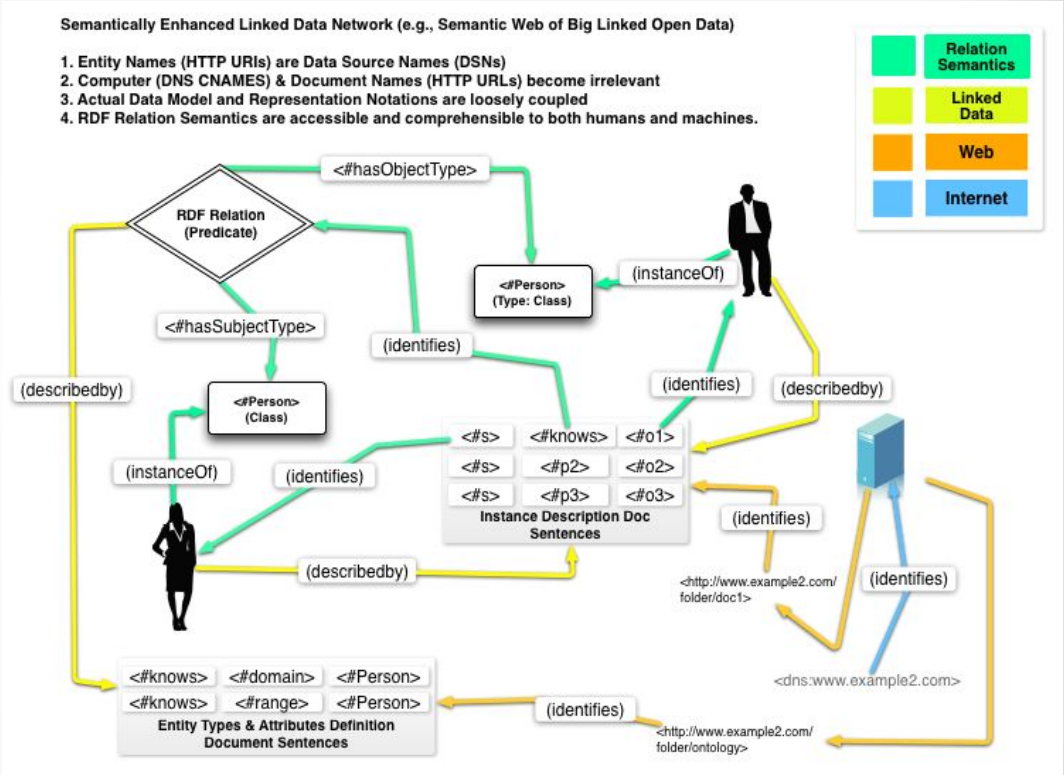
A collection of **Entity Type** and **Entity Relationship Type** definitions associated with a realm of discourse.

Ontologies are loosely-coupled and shareable, if constructed using Linked Data Principles.



How Do You Create an Ontology?

- Name **Entity Types** & **Relationship Types** using a **Hyperlink**
- Describe Entity Types and Relationship Types using **Structured Sentences** where the **Subject, Predicate, and Object** (optionally) are named using a **Hyperlink**
- Save and Publish **Ontology Document**

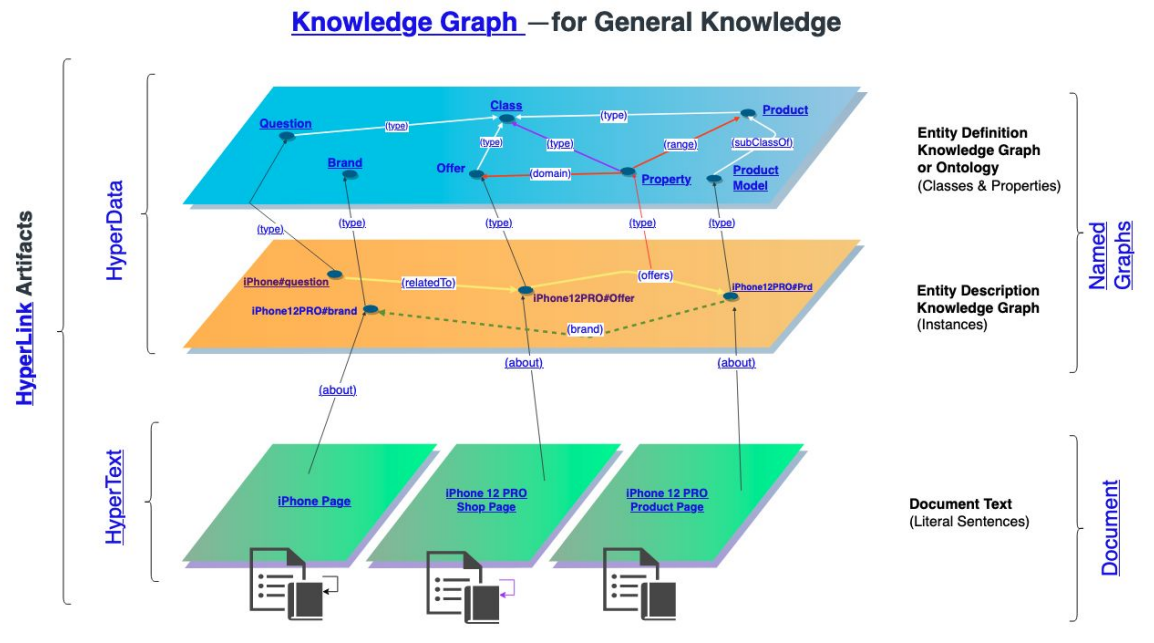


Knowledge Graphs

(Semantic Web Rebranded)

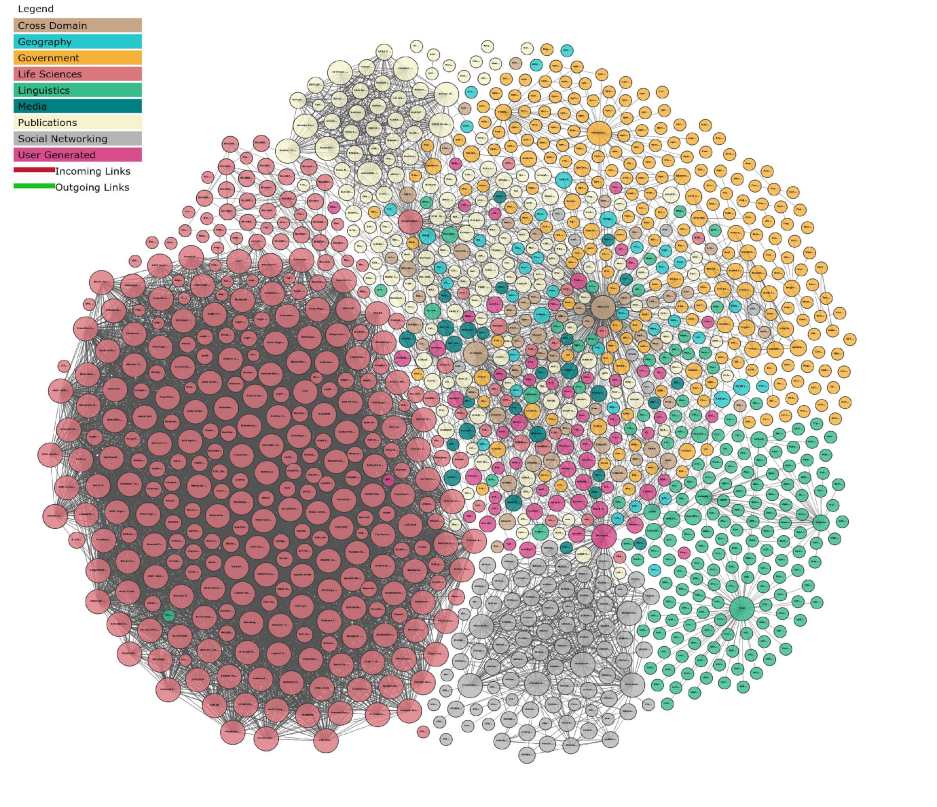
What is a Knowledge Graph?

A collection of **Entities, Entity Types**, and **Entity** Relationship Types that manifests as an intelligible **Web of Data** informed by an **Ontology**



How Do You Create a Knowledge Graph?

- **Name** anything using a **Hyperlink**
- **Describe** everything using **Structured Sentences** where the **Subject, Predicate, and Object** (optionally) are named using a **Hyperlink**



The Semantic Web Vision

&

Journey

Semantic Web Timeline

Year	Event	Key Players
1998	The term "Semantic Web" is coined as a vision of a web where data is structured and linked using common formats and vocabularies	Tim Berners-Lee, James Hendler, and Ora Lassila
2001	The first article on the Semantic Web is published in Scientific American, describing the benefits and challenges of making web data machine-readable	Tim Berners-Lee, James Hendler, and Ora Lassila
2004	The W3C (World Wide Web Consortium) releases the RDF (Resource Description Framework) and OWL (Web Ontology Language) standards for representing and linking data on the web	W3C RDF and OWL Working Groups
2006	The term "Linked Data" is introduced as a set of best practices for publishing and interlinking data on the web using URIs, HTTP, RDF, and SPARQL	Tim Berners-Lee
2007	The Linking Open Data (LOD) project is launched as a community effort to create a web of open data sets that are linked using RDF and SPARQL	LOD Community – Lead By Kingsley Idehen
2008	The first version of DBpedia, a large-scale knowledge base extracted from Wikipedia, is released as one of the core data sets of the LOD cloud	DBpedia Project
2010	Google announces its Knowledge Graph, a system that organizes information from various sources into a graph of entities and their relationships	Google
2012	Schema.org is launched as a collaborative initiative by Google, Microsoft, Yahoo, and Yandex to create and promote a common vocabulary for structured data on the web	Schema.org Consortium
2013	Wikidata is launched as a free and open knowledge base that collects structured data from various sources, such as Wikipedia, Wikimedia Commons, and other external databases	Wikimedia Foundation
2016	The W3C releases the SHACL (Shapes Constraint Language) standard for validating RDF data against a set of rules or constraints	W3C SHACL Working Group

The Vision

Introduced in 2001, via a scientific america article authored by Tim Berners-Lee, Jim Hendler, and Ora Lassila

Web composed of structured data that harnesses the power of logic expressed in natural language sentences.

Challenges

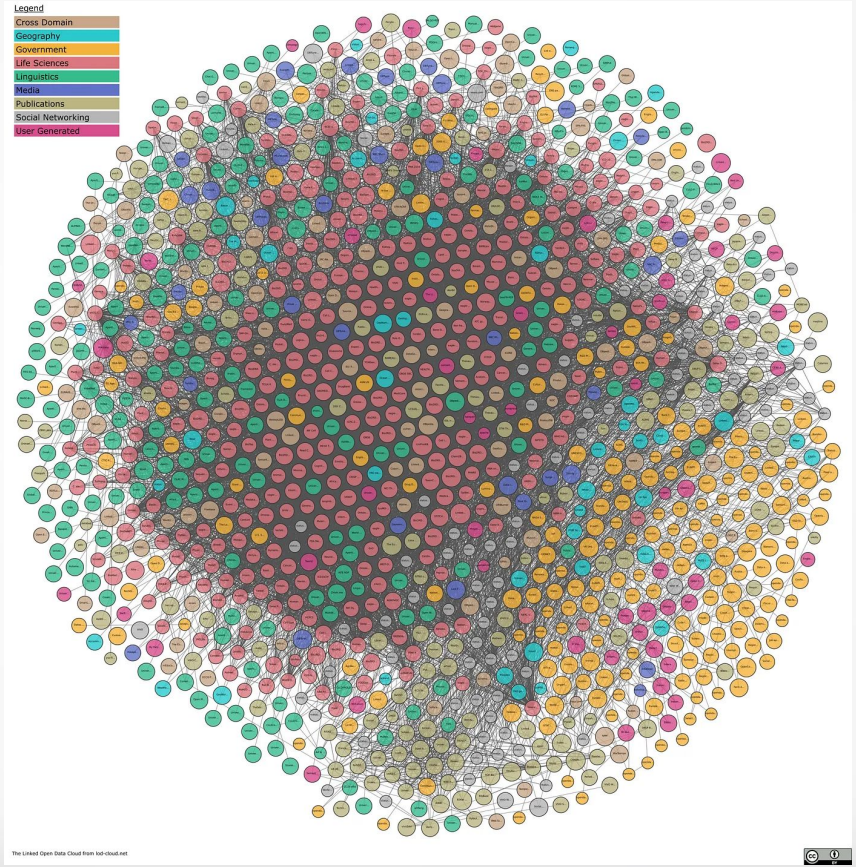
- Fantastic idea plagued by terrible marketing communications.
- Similar to the Web itself, but such a novel concept was challenging for many to comprehend, let alone accept.
- Engaging developers and the emergence of productivity tools also presented a significant challenge.

The Semantic Web Vision Today

Linked Open Data Cloud

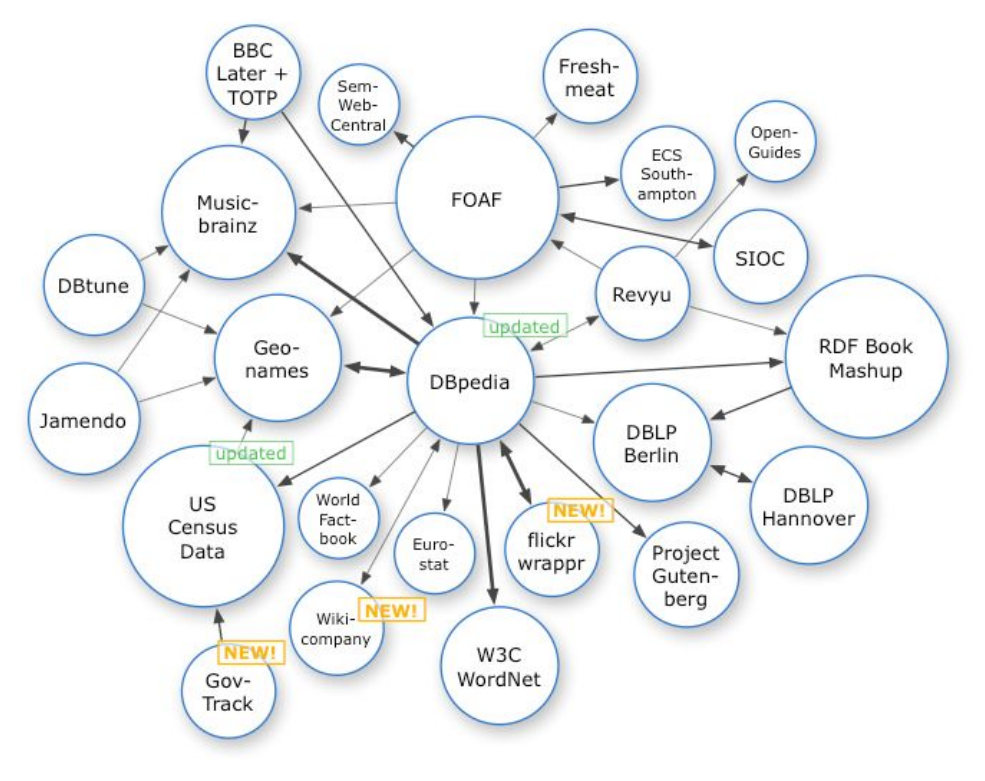
Today, it's a massive collective covering a vast range of mankind's knowledge across a vast number of domains…

It also includes SPARQL Query Service endpoints for powerful querying.



Linked Open Data Cloud

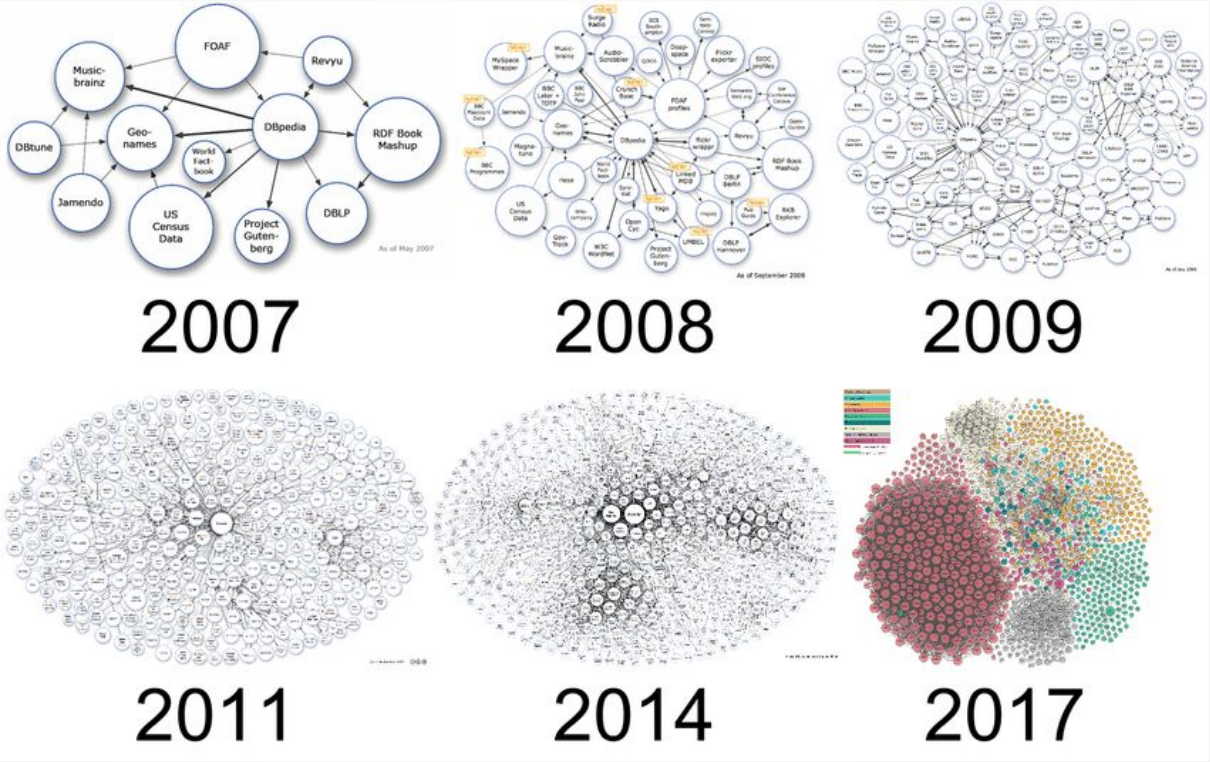
In the beginning circa 2007, courtesy of DBpedia (a Knowledge Graph derived from Wikipedia)



LOD Cloud Version 1.0 Depiction (circa 2007)

Linked Open Data Cloud

Growth over time…



LOD Cloud Growth Depiction

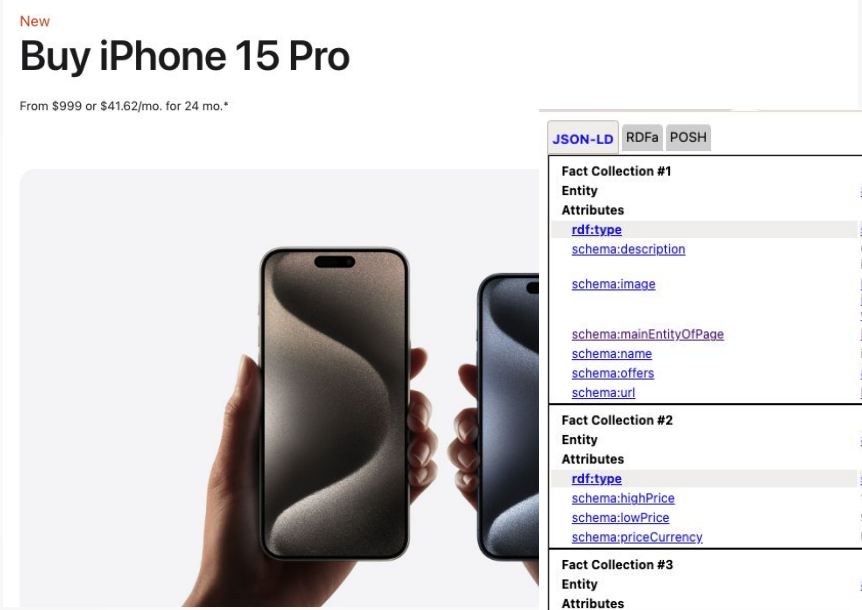
Schema.org and Semantic SEO (SSEO)

Web pages are increasing evolving into Knowledge Graph containers, courtesy of SSEO incentives from Google associated with (E)xpertise, (A)uthority, and (T)rust (E-A-T).

That's why Apple Product Pages also include Offers & FAQs Knowledge Graphs

Semantic SEO Tools

Web Pages are now Structured Data Sources, increasingly on like Tables in a DBMS, but with better access and connectivity due to Linked Data Principles.



JSON-LD | RDFa | POSH

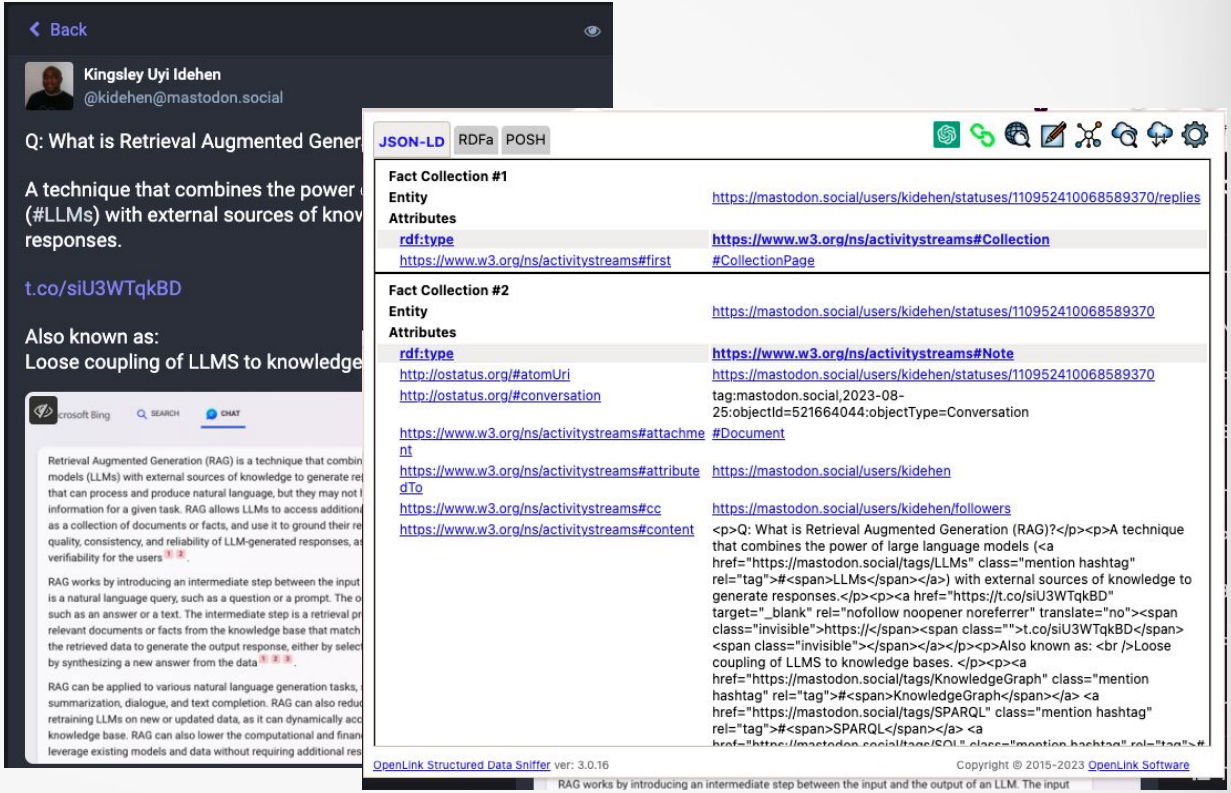
Fact Collection #1	
Entity	#Product
Attributes	
rdf:type	schema:Product
schema:description	Get $40 - $650 off iPhone 15 Pro or iPhone 15 Pro Max when you trade in an iPhone 7 or newer. 0% financing available. Buy now with free shipping.
schema:image	https://store.storeimages.cdn-apple.com/4982/as-images.apple.com/is/iphone-15-pro-model-unselect-gallery-1-202309?wid=2560&hei=1440&fmt=jpeg&qlt=95&.v=1693010533609
schema:mainEntityOfPage	https://www.apple.com/shop/buy-iphone/iphone-15-pro
schema:name	iPhone 15 Pro
schema:offers	#AggregateOffer
schema:url	https://www.apple.com/shop/buy-iphone/iphone-15-pro
Fact Collection #2	
Entity	#AggregateOffer
Attributes	
rdf:type	schema:AggregateOffer
schema:highPrice	1499(xsd:integer)
schema:lowPrice	999(xsd:integer)
schema:priceCurrency	USD
Fact Collection #3	
Entity	#Product_1
Attributes	
rdf:type	schema:Product
schema:description	Get $40 - $650 off iPhone 15 Pro or iPhone 15 Pro Max when you trade in an iPhone 7 or newer. 0% financing available. Buy now with free shipping.
schema:image	https://store.storeimages.cdn-apple.com/4982/as-images.apple.com/is/iphone-15-pro-model-unselect-gallery-1-202309?

OpenLink Structured Data Sniffer ver: 3.0.16 Copyright © 2015-2023 OpenLink Software

OPENLINK SOFTWARE
Making Technology Work For You®

ActivityStreams based Fediverse

Strong desire for decentralized social networks has also opened up another Knowledge Graph frontiers for Micro Blogging (Twitter like services), Blogging, Discussion Forums, and Bookmarking



OPENLINK SOFTWARE
Making Technology Work For You®

Semantic Web Vision Problem

No user-friendly killer app for an application of **Hypertext** aimed at harnessing the power of **Natural Language** via enhanced connectivity unleashed by the **Internet and World Wide Web!**

Semantic Web Vision Problem – Contd

User Interface has been the main stumbling block impeding broader understanding and explosive use of what a Semantic Web offers.

Addressing The User Interface Problem with Large Language Models (LLMs)

Computer User Interfaces: Background

Command Line Interfaces (CLIs): The initial method of interacting with computers

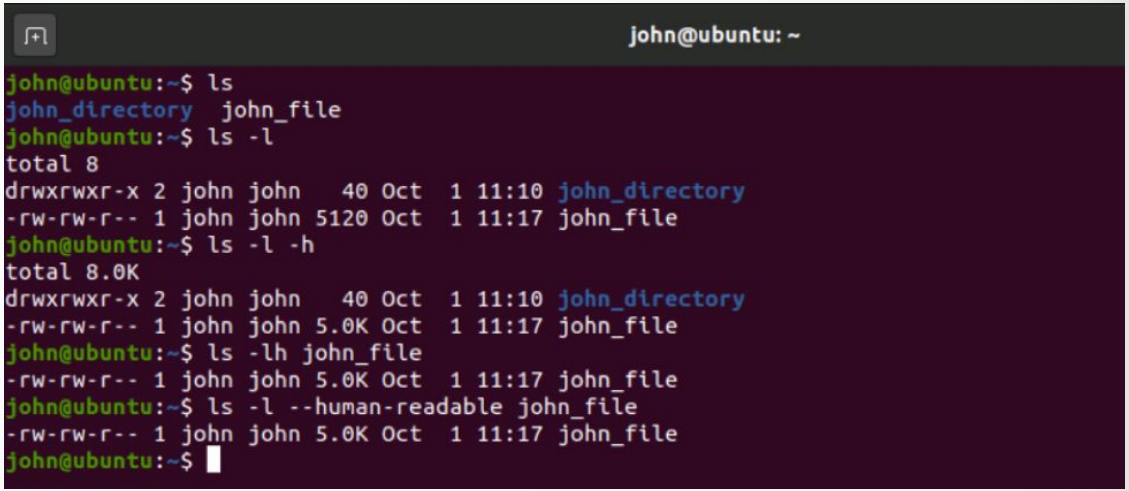
Graphical User Interfaces (GUIs): Derived from widget-based interactions stemming from innovations by Xerox Parc (popularized by Apple and later Microsoft)

CLI Challenges

Traditional command interfaces demand strict syntax adherence.

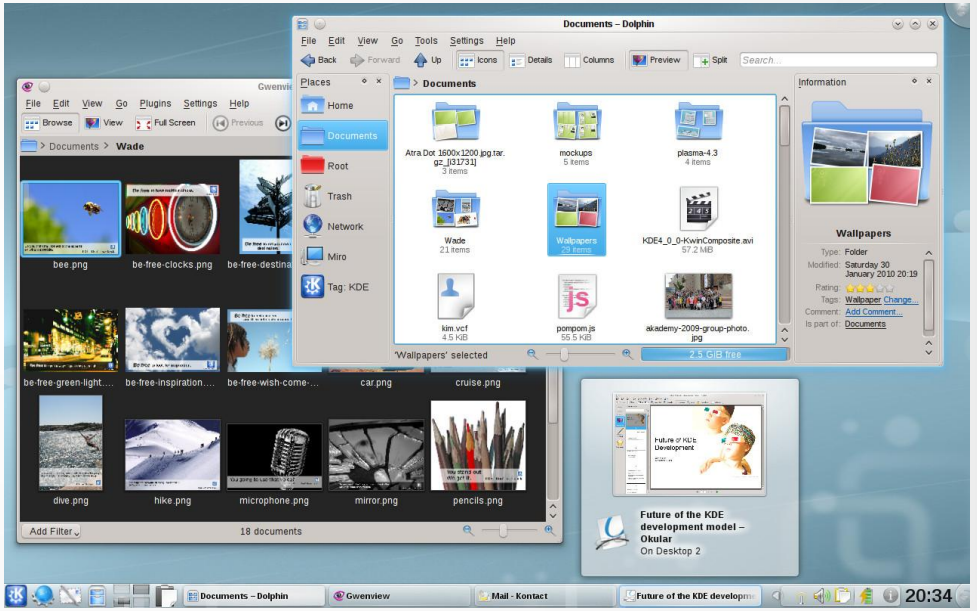
Often face issues due to typographical errors

Conventional documentation can be challenging to navigate.



GUI Challenges

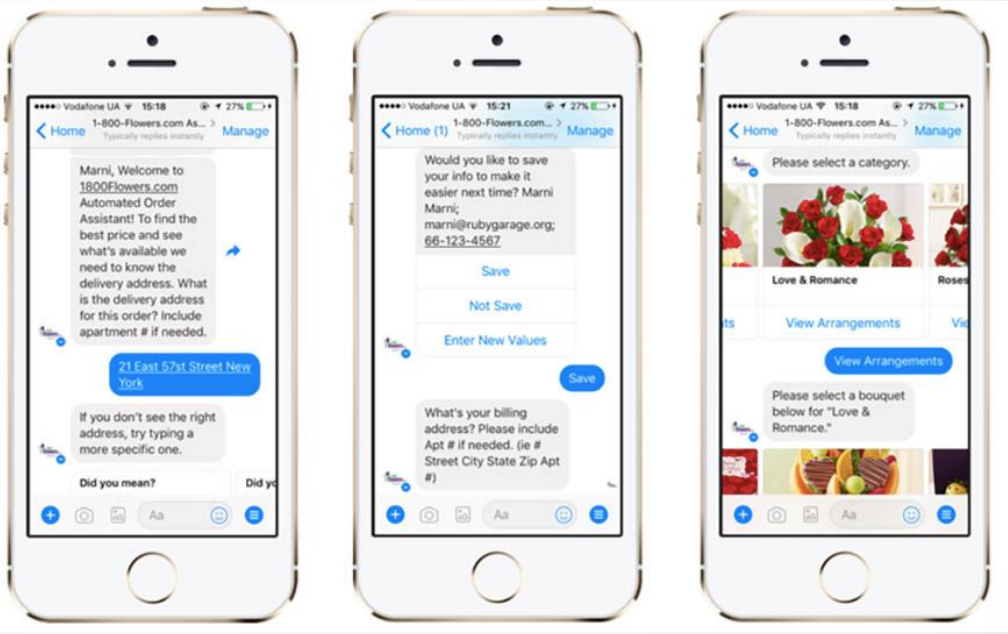
Sentences were replaced by a primitive collection of widgets that merely masquerade as solutions to the underlying interaction issues, eroding productivity and utility in their quest to oversimplify tasks



Primitive Conversational Interfaces

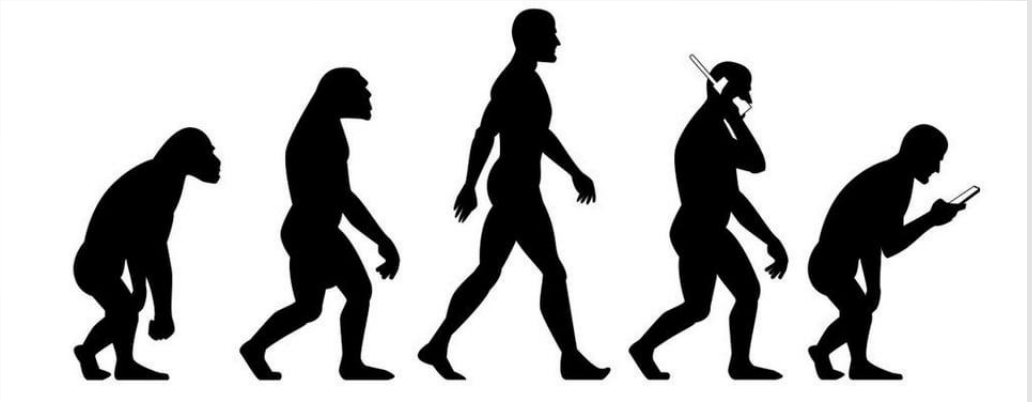
Early attempts to offer an alternative to the CLI and GUI interface challenges that ultimately frustrated users due to an inability to handle sentence variety.

Typical examples include the first generation of support and helpdesk bots.



The Mobile Phone Era

Punctuating the devolutionary impact of not addressing the vital need of natural language processing for computer user interface and interaction is demonstrated in today's mobile phone-dominated era.



LLMs: a Game-Changing Development

Simulated comprehension of sentence structure, syntax, and underlying semantics

Allow a multitude of sentence variations to be accurately mapped to a specific interpretation, which can be associated with specific task execution

LLMs: Challenges

The growth in the size of LLM models has already hit its tipping point [2]

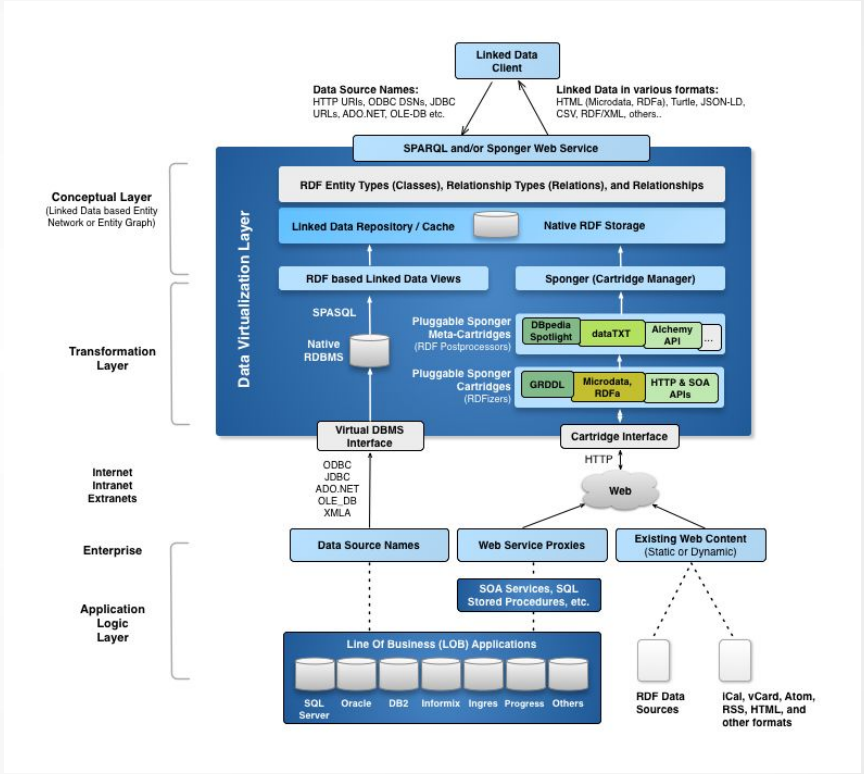
Hallucinations due to the fluid nature of data, information, and knowledge

LLMs should not mistaken for reliable knowledge bases

LLMs & Retrieval Augmented Generation (RAG)

Loose-coupling LLMs with domain-specific knowledge bases provides a powerful and practical strategy for the LLM-hallucination problem.

This boils down to leveraging the code-generating prowess of LLMs, flexibility of declarative query languages (SQL, SPARQL, SPASQL, or GraphQL), and global data connectivity provided by structured data representation using Linked Data Principles.



Semantic Web

&

LLM-based Chatbot Symbiosis

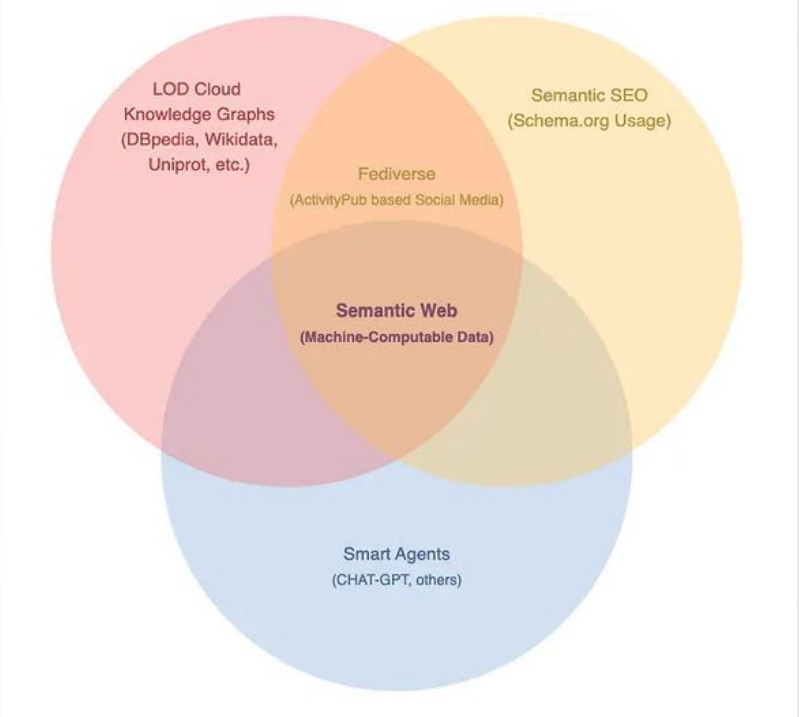
NATURAL LANGUAGE

"Natural Languages are the most sophisticated systems of communication ever developed." – John F. Sowa

"Once you have a truly massive amount of information integrated as knowledge, then the human-software system will be superhuman, in the same sense that mankind with writing is superhuman compared to mankind before writing." – Douglas Lenat

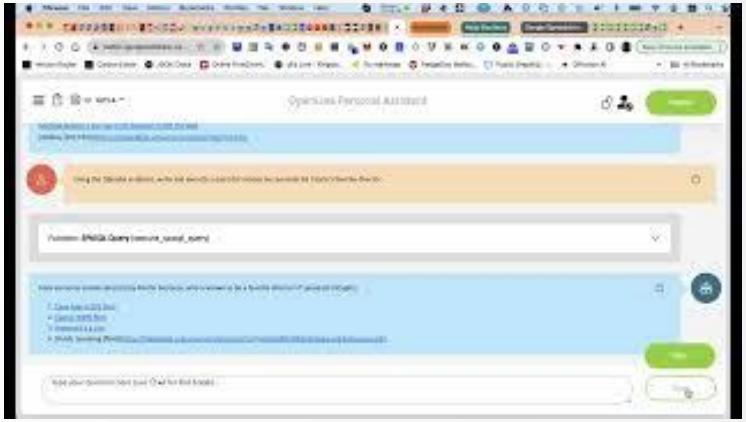
The Best of All Worlds

- Widespread use of LLM-powered natural language processors + rapidly growing Semantic Web (a vast collection of knowledge graphs curated by domain experts)
- LLM hallucinations are significantly mitigated when domain-specific knowledge graphs are loosely integrated
- UI/UX offered by LLM-powered chatbots presents a powerful solution to Semantic Web's core challenges.



Workflow for Optimal Use of LLM-based Bots

1. Identify data sources.
2. Create a virtualization layer for a knowledge graph.
3. Document the layer with HTML.
4. Integrate with LLM bot using SQL or SPARQL.
5. Foster a human-reinforced feedback loop.



Additional Information

- Understanding Data Presentation
- ChatGPT and Semantic Web Symbiosis
- Leveraging LLM-Based Conversational Assistants (Bots) for Enhanced Software Interaction
- Glossary of Terms
- What is the Linked Open Data Cloud, and Why is it Important?
- What is Small Data, and Why is it Important?

Q&A