

# Schema.org

Eduardo Vital Alencar Cunha

# Agenda

- A Web sem conexão
- O Schema.org
- Vantagens & Problemas

# A Web Sem Conexão

# Como receber uma visita

O endereço  
já é  
conhecido

ou

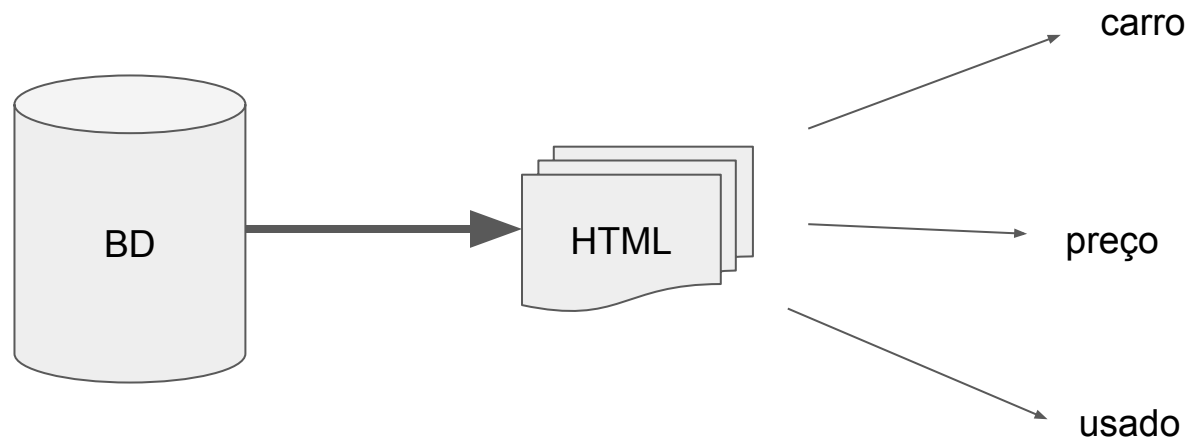
Alguém te  
referencia

Centralizadores de conteúdo



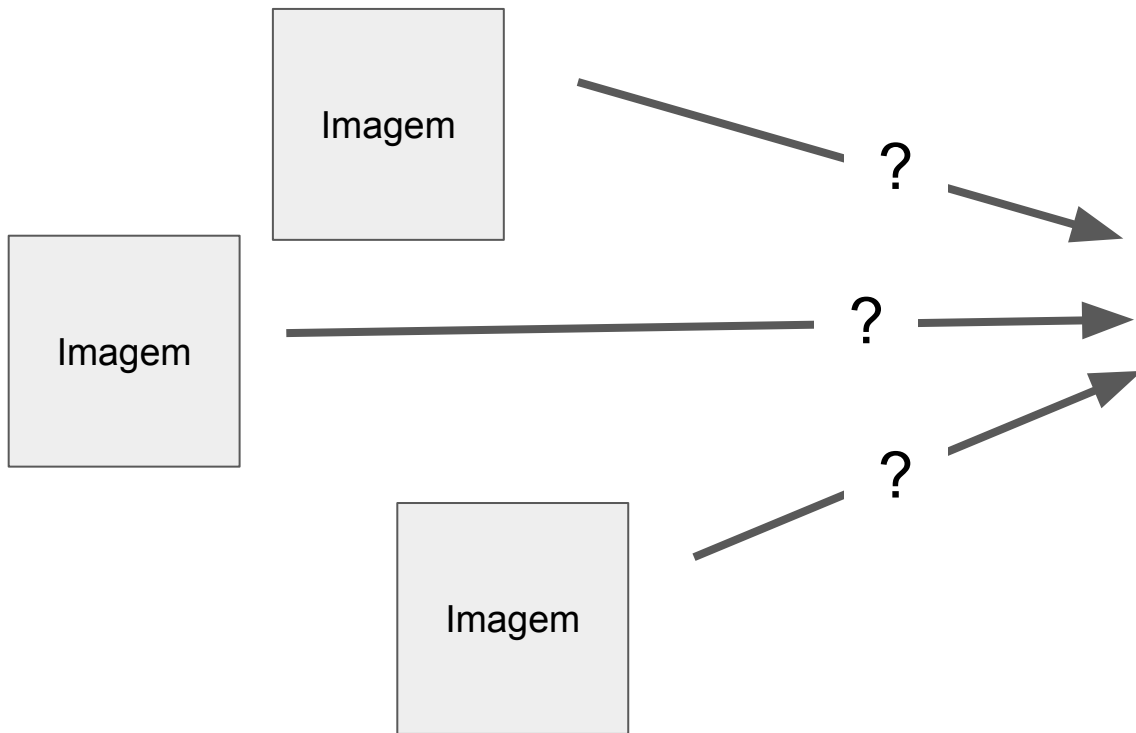
Buscadores

# O processo de extração de dados



Os buscadores baseavam-se em  
palavras-chaves encontradas no  
corpo da página HTML

# O problema...



# Protocolos criados <sup>[1]</sup><sub>[2]</sub>

- **Sitemap** criado pelo Google em **2006** e adotado por todos buscadores progressivamente
- **Rich Snippets** criado pelo Google em **2009**
- **Open Graph Protocol** criado pelo Facebook em **2010**
- **Tiles** criado pelo Bing em **2011**
- **Cards** criado pelo Twitter em **2011**

**Cada buscador recomendava um vocabulário específico**

O Schema.org



# Convergência <sup>[2]</sup>

Em junho de 2011, se uniram para criar o Schema.org:

- Google
- Microsoft
- Yahoo

E em novembro do mesmo ano:

- Yandex

# Schema.org

Uma **coleção de vocabulários** baseado em sintaxes existentes normatizadas [2]

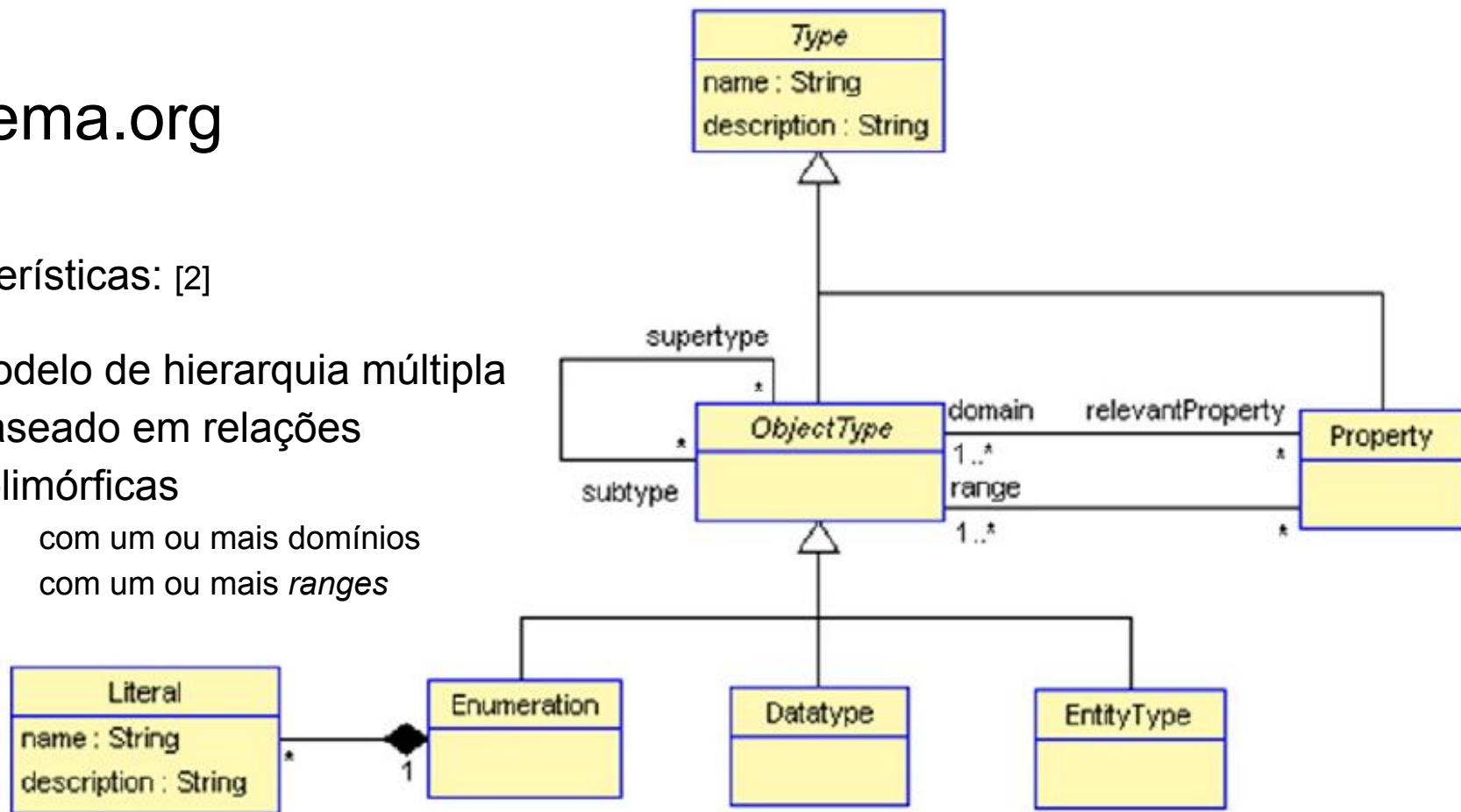
Uma tentativa de definir um amplo vocabulário compartilhado focado em conceitos populares. É uma **ontologia intermediária**. [3]



# Schema.org

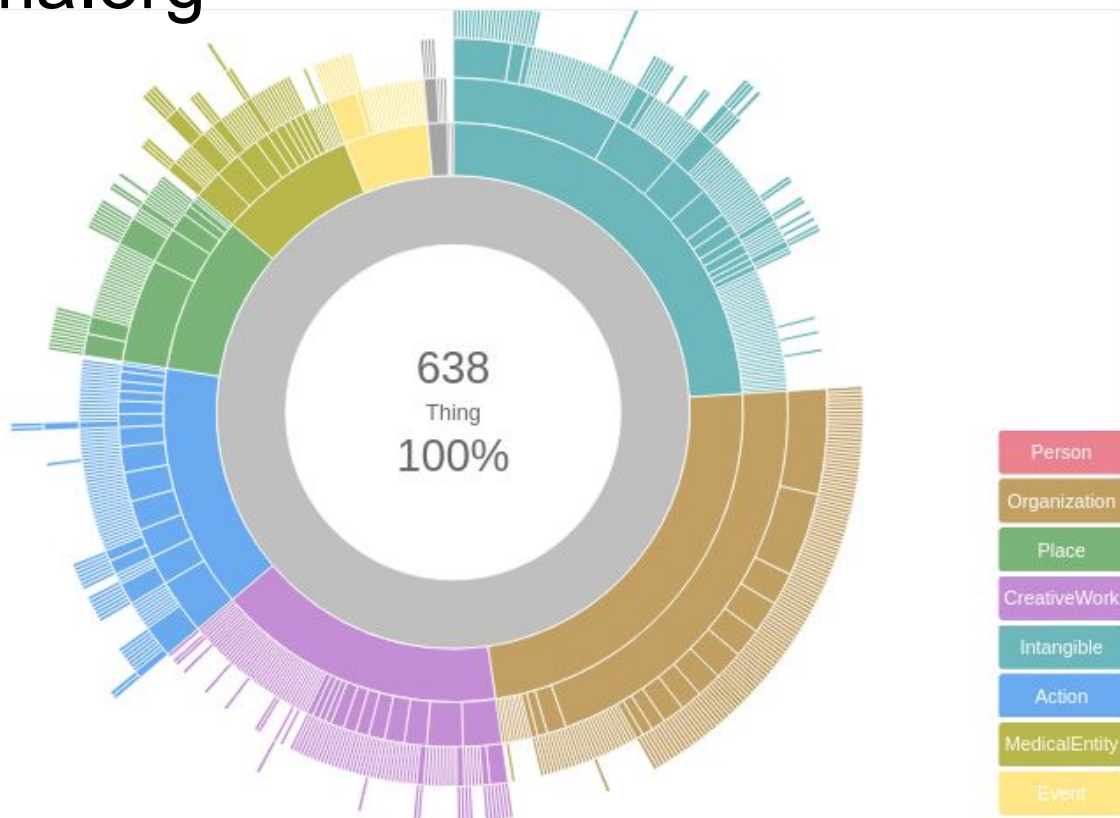
Características: [2]

- Modelo de hierarquia múltipla
- Baseado em relações polimórficas
  - com um ou mais domínios
  - com um ou mais *ranges*



Representação em UML do metamodelo do Schema.org [5]

# Schema.org



**297 classes**  
**187 relações**  
em 2011 [2]

**638 classes**  
**965 relações**  
em 2017

Representação Hierárquica em formato *SunBurst* do Schema.org [4]

# Impacto

- **3,2%** da web utilizava o Schema.org em 2012 [1]
- **14,5%** da web utilizava o Schema.org em 2014 [1]
- **31,3%** de uma amostra de 10 bilhões de páginas utilizava o Schema.org em 2016 [2]
- Em média, cada página que utiliza o Schema.org referência 6 entidades, realizando 26 declarações lógicas entre si [2]
- É o único vocabulário utilizado por mais de  $\frac{1}{4}$  das páginas encontradas em grandes buscadores [2]

# Impacto

Category	Sites
News	nytimes.com, guardian.com, bbc.co.uk
Movies	imdb.com, rottentomatoes.com, movies.com
Jobs / Careers	careerjet.com, monster.com, indeed.com
People	linkedin.com, pinterest.com, familysearch.org, archives.com
Products	ebay.com, alibaba.com, sears.com, cafepress.com, sulit.com, fotolia.com
Video	youtube.com, dailymotion.com, frequency.com, vinebox.com
Medical	cvs.com, drugs.com
Local	yelp.com, allmenus.com, urbanspoon.com
Events	wherevent.com, meetup.com, zillow.com, eventful.com
Music	last.fm, myspace.com, soundcloud.com

Utilizadores do Schema.org [7]

# Vantagens & Problemas

# Vantagens

- Facilidade de publicadores de conteúdo e desenvolvedores adotarem
- Provê diversos exemplos de utilização em sua documentação
- Complexidade é adicionada progressivamente
- Permite extensões (48 ontologias usa o Schema.org como base). [6]
- Colaboração aberta e transparente via GitHub, com muitas modificações propostas por agentes externos



# Problemas do Schema.org

- Não é a solução final para vocabulário da Web Semântica
  - há muitos vocabulários que abordam escopos mais específicos e detalhados
- O Schema.org precisa enfrentar os desafios da escalabilidade a medida que cresce o número de extensões que aceita
  - ambiguidade
  - complexidade
- É difícil para o usuário decidir quais vocabulários utilizar em meio a tantos
- Ainda é custoso utilizar o Schema.org para marcar todo o conteúdo relevante de uma página web

# Referências Bibliográficas

1. MIKA, Peter. **On Schema. org and why it matters for the web**. IEEE Internet Computing, v. 19, n. 4, p. 52-55, 2015.
2. GUHA, Ramanathan V.; BRICKLEY, Dan; MACBETH, Steve. **Schema.org: evolution of structured data on the web**. Communications of the ACM, v. 59, n. 2, p. 44-51, 2016.
3. RONALLO, Jason. **HTML5 Microdata and Schema.org**. Code4Lib Journal, v. 16, 2012.
4. BRICKLEY, Dan. **Schema.org hierarchy Sunburst**. <http://bl.ocks.org/danbri/1c121ea8bd2189cf411c>, 2016.
5. TORT, Albert; OLIVÉ, Antoni. **An approach to website schema.org design**. Data & knowledge engineering, v. 99, p. 3-16, 2015.
6. MCCARTHY, Matthew T. **The Semantic Web and Its Entanglements**. Science, Technology and Society, v. 22, n. 1, p. 21-37, 2017.
7. NOGALES, Alberto et al. **Linking from Schema. org microdata to the Web of Linked Data: An empirical assessment**. Computer Standards & Interfaces, v. 45, p. 90-99, 2016.