

第七章 描述逻辑和语义网

戴洪良

计算机科学与技术学院/人工智能学院

hongldai@nuaa.edu.cn



描述逻辑简介

- 描述逻辑指一系列表示概念、它们的属性、及它们间关系的形式化知识表示语言。
- 表达能力通常强于命题逻辑，但弱于谓词逻辑，是谓词逻辑的可判定 (decidable) 子集
- 适合应用于对本体的表示和推理
- 是语义网本体语言OWL (Web Ontology Language) 的基础

为什么使用描述逻辑

- 方法的两个角度：科学性和工程性
 - 科学性：形式化，详细地分析问题
 - 工程性：快速高效地完成任务
- 描述逻辑可以较好地平衡语言的表达能力和推理的复杂程度
 - 有多种不同的描述逻辑语言：ALC、SHIQ、SROIQ等，可根据具体应用需要选择

概念、角色、常元

- 概念、角色、常元是描述逻辑中用于构建句子的基本元素
 - 描述逻辑中的句子也可被赋予真值 (True, False)
- 概念 (Concept) : 对应类别, 与谓词逻辑中的一元谓词一致
 - 如: Person, Student, Female
- 角色 (Role) : 对应关系, 与谓词逻辑中的二元谓词一致
 - 如: AgeOf, ParentOf, AreaOfStudy
- 常元 (Constant) : 指代个体, 与谓词逻辑中的个体常元一致
 - 如: John, Smith

TBox、ABox

- 描述逻辑中的KB包含两部分：TBox、ABox
- TBox: 术语定义 (Define terminology) ; 包含描述概念层次、概念间关系的句子
 - 如: $Mother \sqsubseteq Female, Person \equiv Human$
- ABox: 断言 (Give assertions) ; 包含关于个体的知识, 包括个体与概念的关系, 个体间的关系
 - 如: $Father(John), ParentOf(John, Bob)$

ALC

- ALC是最基本的描述逻辑语言

ALC

- 其KB包括TBox、ABox
- 符号：
 - 概念构造算子：合取 (\sqcap)、析取 (\sqcup)、非 (\neg)
 - 量词约束：全称量词 (\forall)、存在量词 (\exists)
 - 顶概念 (top concept) 和底概念 (bottom concept)
 - 分别为 \top , \perp
- 概念断言符号：包含 (\sqsubseteq)、等价 (\equiv)

ALC语言

- 基于已有概念得到新的概念
- 设 C 和 D 是概念, R 是角色
 - $\neg C$ 表示概念: 不是 C 的事物
 - $C \sqcap D$ 表示概念: 既是 C 又是 D 的事物
 - 如: $Female \sqcap Human$
 - $C \sqcup D$ 表示概念: 是 C 或是 D 或既是 C 又是 D 的事物
 - 如: $Female \sqcup Male$
 - $\forall R.C$ 表示概念: 所有与其有 R 关系的事物都是 C 的事物
 - 如: $\forall ParentOf.Female$: 其孩子都是女性的事物
 - $\exists R.C$ 表示概念: 一些与其有 R 关系的事物是 C 的事物
 - 如: $\exists ParentOf.Female$: 有孩子是女性的事物

ALC知识库

- TBox中的句子
 - 包含 (subsumption) : $C \sqsubseteq D$, 属于概念C的个体都属于D
 - 等价 (equivalence) : $C \equiv D$
- ABox中的句子
 - $C(a)$, 其中C是概念, a 是常元
 - $R(a, b)$, 其中R是角色, a 和 b 是常元

顶概念和底概念

- 顶概念 \top
 - 每个个体都属于它
 - 如: $\top \sqsubseteq Male \sqcup Female$
- 底概念
 - 没有个体属于它
 - 如: $Male \sqcap Female \sqsubseteq \perp$

ALC知识库 – 例

- TBox:

$Person \sqsubseteq Animal \sqcap Biped$

$Woman \equiv Person \sqcap Female$

$Mother \equiv Woman \sqcap \exists ParentOf. Person$

$Parent \equiv Mother \sqcup Father$

是某些人的家长

$Man \equiv Person \sqcap \neg Woman$

$MotherWithoutDaughter \equiv Mother \sqcap \forall ParentOf. \neg Female$

$GrandMother \equiv Woman \sqcap \exists ParentOf. Parent$

其所有孩子不是女性

其一些孩子是家长

- ABox:

$GrandMother(Sally)$

$(Person \sqcap Male)(John)$

解释 (Interpretation)

- ALC的一个解释可记为 $\mathcal{I} = \langle \Delta, \cdot^{\mathcal{I}} \rangle$
 - 论域 Δ : 非空的对象集
 - 解释函数 $\cdot^{\mathcal{I}}$: 将结构映射到论域
- $\cdot^{\mathcal{I}}$ 将每个概念名A映射到一个论域的子集 $A^{\mathcal{I}} \subseteq \Delta$
- $\cdot^{\mathcal{I}}$ 将每个角色名R映射到二元关系 $R^{\mathcal{I}} \subseteq \Delta \times \Delta$
- $\cdot^{\mathcal{I}}$ 将每个常元映射到一个对象 $a^{\mathcal{I}} \in \Delta$
- $\top^{\mathcal{I}} = \Delta$ $\perp^{\mathcal{I}} = \emptyset$

解释 (Interpretation)

- 设C、D是概念，R是角色

- $(\neg C)^{\mathcal{I}} = \Delta \setminus C^{\mathcal{I}}$
- $(C \sqcap D)^{\mathcal{I}} = C^{\mathcal{I}} \cap D^{\mathcal{I}}$
- $(C \sqcup D)^{\mathcal{I}} = C^{\mathcal{I}} \cup D^{\mathcal{I}}$
- $(\forall R.C)^{\mathcal{I}} = \{x \mid y \in C^{\mathcal{I}} \text{ for every } y \text{ s.t. } (x, y) \in R^{\mathcal{I}}\}$
- $(\exists R.C)^{\mathcal{I}} = \{x \mid y \in C^{\mathcal{I}} \text{ for some } y \text{ s.t. } (x, y) \in R^{\mathcal{I}}\}$

$$(\forall \text{parentOf.female})^{\mathcal{I}} = \{x \mid \text{任意满足 } (x, y) \in \text{parentOf}^{\mathcal{I}} \text{ 的 } y, \text{ 都有 } y \in \text{female}^{\mathcal{I}}\}$$

$$(\exists \text{parentOf.female})^{\mathcal{I}} = \{x \mid \text{存在满足 } (x, y) \in \text{parentOf}^{\mathcal{I}} \text{ 的 } y, y \in \text{female}^{\mathcal{I}}\}$$

解释 (Interpretation) - 例

$$\neg Female \sqcap \exists haschild.Human$$

- 解释 $\mathcal{I} = \langle \Delta, \cdot^{\mathcal{I}} \rangle$ $\Delta = \{John, Stan\}$
- $Human^{\mathcal{I}} = \{John, Stan\}$
- $Female^{\mathcal{I}} = \{\}$
- $hasChild^{\mathcal{I}} = \{(John, Stan)\}$
- $(\neg Female)^{\mathcal{I}} = \Delta - Female^{\mathcal{I}} = \{John, Stan\}$
- $(\exists hasChild.Human)^{\mathcal{I}} = \{John\}$
- $(\neg Female \sqcap \exists hasChild.Human)^{\mathcal{I}} = \{John\}$

解释

- $C \sqsubseteq D$ 在 \mathcal{I} 中为真当且仅当 $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$
- $C \equiv D$ 在 \mathcal{I} 中为真当且仅当 $C^{\mathcal{I}} = D^{\mathcal{I}}$
- $C(a)$ 在 \mathcal{I} 中为真当且仅当 $a^{\mathcal{I}} \in C^{\mathcal{I}}$
- $R(a, b)$ 在 \mathcal{I} 中为真当且仅当 $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in R^{\mathcal{I}}$

ALC中的推理

给定：

- 一个KB (a set of sentences)
- 一个句子 α

问： KB是否语义蕴含 α ? $KB \models \alpha$?

即： 所有令KB为真的解释是否也令 α 为真

ALC中推理的类型

- 实例检测
 - 如: $KB \models C(a) \quad KB \models R(a, b)$
- 包含检测
 - 如: $KB \models C \sqsubseteq D$
- 等价检测
 - 如: $KB \models C \equiv D$

转化为可满足性问题

- 实例检测

$KB \models C(a)$ 当且仅当 $KB \cup \{\neg C(a)\}$ 不可满足

- 包含检测

$KB \models C \sqsubseteq D$ 当且仅当 $KB \cup \{(C \sqcap \neg D)(b)\}$ 不可满足

- 等价检测

$KB \models C \equiv D$ 当且仅当 $KB \cup \{(C \sqcup D)(b), (\neg C \sqcup \neg D)(b)\}$ 不可满足

Tableau算法

- 将概念转化为Negational Normal Form (NNF)
 - 非 (\neg) 只出现在概念符号前, 如 $\neg(C_1 \sqcap C_2) \equiv \neg C_1 \sqcup \neg C_2$
- 构建初始的完备图 (completion graph)
 - 图的节点表示一个个体, 并标注了它需满足的概念集合
 - 图的边标注了对应两个个体需满足的角色集合
- 应用规则对图进行扩展, 直到产生矛盾 (不可满足) 或无法再扩展 (可满足)
 - 每次扩展将一些原本隐式蕴含的约束显示表示出来

该算法可靠且完备

描述逻辑的难点

- 构建难度大
 - 人工或自动标注都无法保证数据的准确性
- 错误定位难
 - 实践中缺乏有效的错误定位手段

语义网 (Semantic Web)

World Wide Web

- 万维网由英国计算机科学家Tim Berners-Lee于1989年发明，1991年向世界公开
 - 万维网上的资源由网站服务器发布，通过HTTP协议传输，用户通过网页浏览器接收
- 1994年，Berners-Lee在第一届International WWW Conference上展示了语义网（Semantic Web）的想法

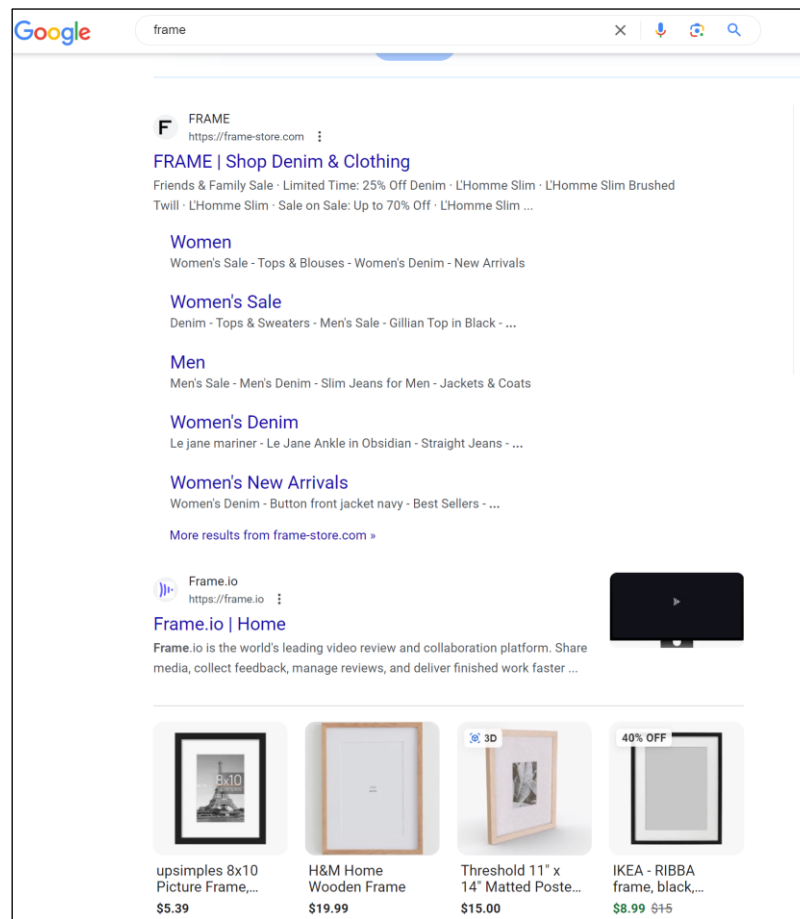
WWW (The Web)

- Web 1.0
 - 获取信息、读信息
 - 大致从1989到2004
 - 网站是静态内容；少量的内容创作者，大部分人只从网上读取信息
 - 能由用户产生内容的方式很少：聊天室、论坛
- Web 2.0
 - 读写信息；用户参与创作、与其他用户交互
 - B站、抖音、Youtube、Facebook等
- Web 3.0
 - 区块链技术、去中心化、元宇宙？
 - 语义网（Semantic Web）（曾）是人们对Web 3.0的愿想之一



当前网络的不足

- 网上的内容是人类易理解的，但计算机难以“理解”
 - HTML主要定义展示时的样式，不描述内容的语义
- 导致在一些应用上出现困难
 - 查找或抽取相关信息
 - 如：“Frame”有多种含义
 - 对信息组合或重用



语义网 (Semantic Web)

- 对the Web进行扩展，使网上的数据变得机器可读 (machine-readable)
 - 是扩展，而不是取代
- 为网上的数据提供语义层次的定义，并将它们连接起来
- 从而使数据的发现、整合和重用变得更高效

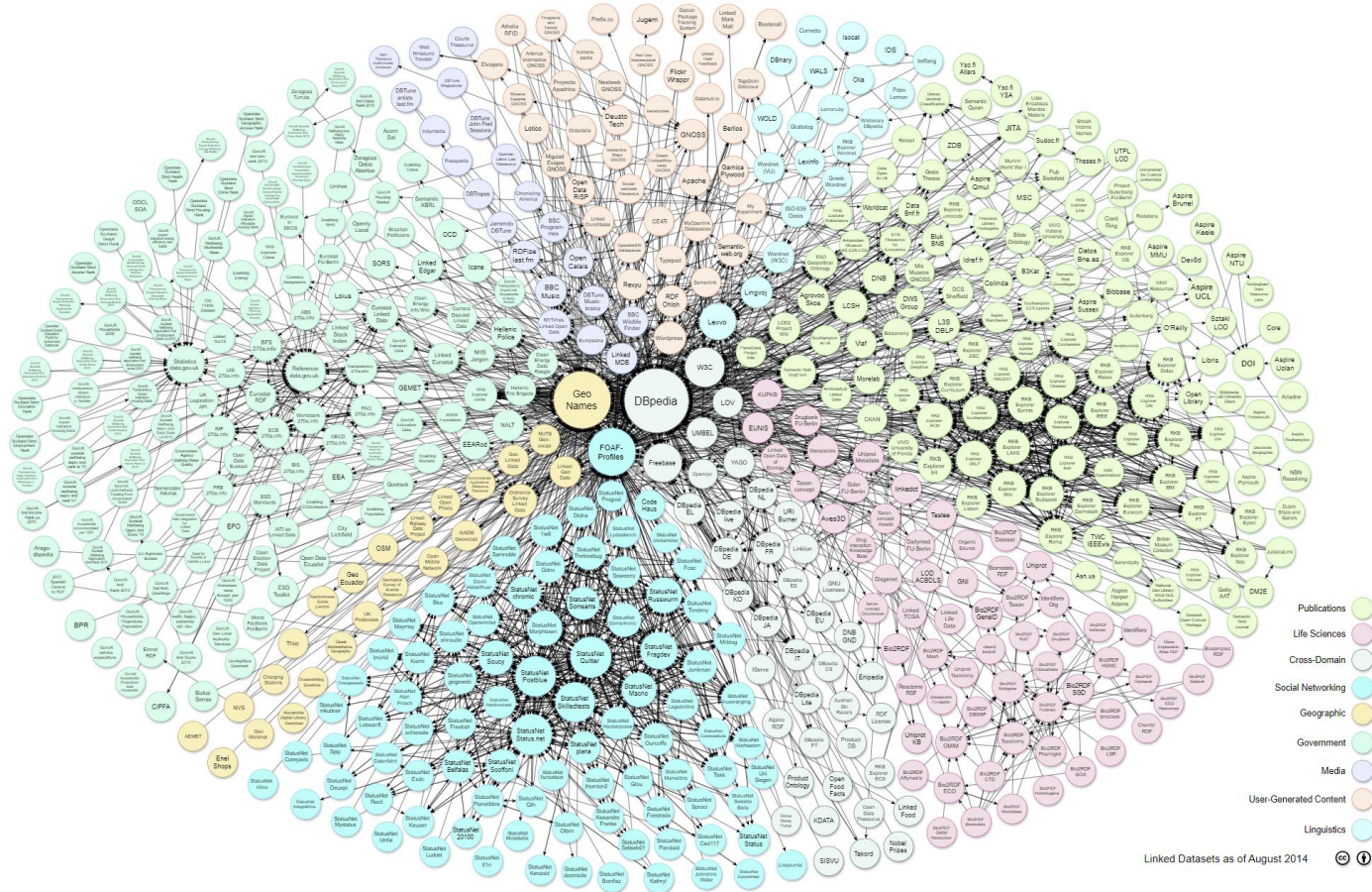
语义网的主要内容

- 关联开放数据 (Linked Open Data)
 - 构建全球的开放互联的数据网络
- 语义标注
 - 当前Web上网页HTML主要用于规定样式，而非标记内容语义
 - 标注网页内容的语义，并使它们与关联开放数据中的相关知识建立联系
 - 可标注的网页内容：文本中的词、短语；图片；文档等。

关联开放数据 (Linked Open Data)

- 开放互联的全球数据网络
- 打破不同组织机构或个人的数据之间的阻隔；方便数据的获取和整合。

<https://lod-cloud.net/>



关联开放数据

- 构建机器可读的，可关联的数据的前提：
 - 1) 对所有资源（人、书、电影、概念）赋予唯一的标识
 - URI
 - 2) 一个统一的对知识进行表示和解释的体系
 - W3C (World Wide Web Consortium) 组织建立了一些标准和推荐
 - W3C是Web标准的主要制定发布组织
 - XML、RDF、OWL等

关联开放数据 (Linked Open Data)

- 关联数据的4条准则：
 - 使用统一资源标识符 (URI) 命名事物
 - 使用HTTP URI, 从而使这些名字可以被查找
 - 当URI被查找时, 提供有用的信息
 - 提供其他URI链接, 从而让人能发现更多相关事物
- 数据类型：
 - 本体 (Ontology)
 - OWL
 - 关于实体的事实性数据
 - RDF

统一资源标识符 (URI)

- URI是用于唯一标识一个抽象或物理资源的字符串
 - 可标识网页、书、人、地点、概念等
 - 如： <https://hldai.github.io/index.html>, isbn:9787536692930
- 与URL的区别：
 - URL标识了一个资源的地址，可通过该地址访问该资源，但用URI不一定能访问到对应资源
 - URL是URI的一种

RDF

- RDF (Resource Description Framework) 是W3C建立的信息表示框架，是在Web上进行数据交换的一种标准数据模型
- RDF采用(*subject, predicate, object*)三元组的形式表示关于资源的信息，这样的三元组称为一个陈述 (statement)
 - 含义为subject和object之间存在由predicate所表示的关系
 - 基于这样的三元组可以构成有向图

RDF – 例

There is a person identified by <http://www.w3.org/People/EM/contact#me>, whose name is Eric Miller, whose title is Dr.

相关predicate的URI:

"whose name is": <http://www.w3.org/2000/10/swap/pim/contact#fullName>
"whose title is": <http://www.w3.org/2000/10/swap/pim/contact#personalTitle>
type: <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>

Person的URI: <<http://www.w3.org/2000/10/swap/pim/contact#Person>>

RDF, 三元组形式表示:

```
<http://www.w3.org/People/EM/contact#me> <http://www.w3.org/2000/10/swap/pim/contact#fullName> "Eric Miller" .  
<http://www.w3.org/People/EM/contact#me> <http://www.w3.org/2000/10/swap/pim/contact#personalTitle> "Dr." .  
<http://www.w3.org/People/EM/contact#me> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>  
                                     <http://www.w3.org/2000/10/swap/pim/contact#Person> .
```


RDF – 例

There is a person identified by <http://www.w3.org/People/EM/contact#me>, whose name is Eric Miller, whose title is Dr.

相关predicate的URI:

"whose name is": <http://www.w3.org/2000/10/swap/pim/contact#fullName>
"whose title is": <http://www.w3.org/2000/10/swap/pim/contact#personalTitle>
type: <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>

Person的URI: <<http://www.w3.org/2000/10/swap/pim/contact#Person>>

RDF, Turtle表示:

```
@prefix eric:    <http://www.w3.org/People/EM/contact#> .  
@prefix contact: <http://www.w3.org/2000/10/swap/pim/contact#> .  
@prefix rdf:     <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .  
  
eric:me contact:fullName "Eric Miller" .  
eric:me contact:personalTitle "Dr." .  
eric:me rdf:type contact:Person .
```

RDF – 例

RDF, RDF/XML表示:

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF xmlns:contact="http://www.w3.org/2000/10/swap/pim/contact#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <rdf:Description rdf:about="http://www.w3.org/People/EM/contact#me">
    <contact:fullName>Eric Miller</contact:fullName>
    <contact:personalTitle>Dr.</contact:personalTitle>
    <rdf:type rdf:resource="http://www.w3.org/2000/10/swap/pim/contact#Person"/>
  </rdf:Description>
</rdf:RDF>
```

`rdf:Description`元素中包含由`rdf:about`标识的资源（对应subject）的描述
`contact:fullName`、`contact:personalTitle`、`rdf:type`都是谓词（predicate）
`rdf:resource`标识了object的URI

例 - DBpedia

- DBpedia是通过从Wikipedia抽取结构化数据形成的知识图谱
- 其知识主要从不同语言的Wikipedia的infobox获取，用RDF表示

<https://www.dbpedia.org/>

<https://lookup.dbpedia.org/>

Barbie^[a] is a 2023 **fantasy comedy** film directed by **Greta Gerwig** from a screenplay she wrote with **Noah Baumbach**. Based on the **eponymous fashion dolls** by **Mattel**, it is the first live-action Barbie film after numerous **animated films and specials**. It stars **Margot Robbie** as the title character and **Ryan Gosling** as **Ken**, and follows them on a journey of self-discovery through both Barbieland and the real world following an **existential crisis**. It is also a commentary regarding **patriarchy** and the effects of **feminism**. The supporting cast includes **America Ferrera**, **Michael Cera**, **Kate McKinnon**, **Issa Rae**, **Rhea Perlman**, and **Will Ferrell**.

A live-action Barbie film was announced in September 2009 by **Universal Pictures** with **Laurence Mark** producing. Development began in April 2014, when **Sony Pictures** acquired the **film rights**. Following multiple writer and director changes and the casting of **Amy Schumer** and later **Anne Hathaway** as Barbie, the rights were transferred to **Warner Bros. Pictures** in October 2018. Robbie was cast in 2019, after **Gal Gadot** turned down the role due to scheduling conflicts, and Gerwig was announced as director and co-writer with Baumbach in 2020. The rest of the cast was announced in early 2022. **Principal photography** occurred primarily at **Warner Bros. Studios, Leavesden**, England, and at the **Venice Beach Skatepark** in Los Angeles from March to July 2022.

Barbie premiered at the **Shrine Auditorium** in Los Angeles on July 9, 2023, and was released in the United States on July 21. The film has grossed \$1.446 billion and achieved several milestones, including the **highest-grossing film of 2023** and the **14th highest-grossing film of all time**. Named one of the top 10 films of 2023 by the **National Board of Review** and the **American Film Institute**, it received critical acclaim and **other accolades**, including eight **Academy Award** nominations, winning **Best Original Song** for "**What Was I Made For?**"; the song also won **Golden Globe Award for Best Original Song** while *Barbie* was awarded **Golden Globe Award for Cinematic and Box Office Achievement**.

Plot

Barbie ("Stereotypical Barbie") and fellow dolls reside in Barbieland, a **matriarchal** society populated by different versions of Barbies, **Kens**, and a group of discontinued models who are treated like outcasts due to their unconventional traits. While the Kens spend their days playing at the beach, considering it their



Theatrical release poster

Directed by	Greta Gerwig
Written by	Greta Gerwig Noah Baumbach
Based on	Barbie by Mattel
Produced by	David Heyman Margot Robbie Tom Ackerley Robbie Brenner
Starring	Margot Robbie Ryan Gosling America Ferrera Kate McKinnon

SPARQL

- 对RDF数据进行查询的语言

例：DBpedia提供了在线SPARQL查询功能

查找张艺谋导演的电影：

```
SELECT ?film, ?title
WHERE
{
  ?film dbo:director dbr:Zhang_Yimou .
  ?film rdfs:label ?title .
  FILTER ( LANG ( ?title ) = 'zh' )
}
```

<https://dbpedia.org/sparql>

结果：

SPARQL | HTML5 table

film	title
http://dbpedia.org/resource/House_of_Flying_Daggers	"十面埋伏 (2004年电影)"@zh
http://dbpedia.org/resource/Ju_Dou	"菊豆"@zh
http://dbpedia.org/resource/Riding_Alone_for_Thousands_of_Miles	"千里走单骑"@zh
http://dbpedia.org/resource/Under_the_Hawthorn_Tree_(film)	"山楂树之恋 (电影)"@zh
http://dbpedia.org/resource/Coming_Home_(2014_film)	"归来 (电影)"@zh
http://dbpedia.org/resource/Cliff_Walkers	"悬崖之上"@zh
http://dbpedia.org/resource/Codename_Cougar	"代号美洲豹"@zh
http://dbpedia.org/resource/The_Road_Home_(1999_film)	"我的父亲母亲"@zh
http://dbpedia.org/resource/The_Story_of_Qiu_Ju	"秋菊打官司"@zh
http://dbpedia.org/resource/Sniper_(2022_film)	"狙击手 (电影)"@zh

语义标注

- 标注网页内容的语义，使它们与关联开放数据中的相关知识建立联系
 - 可标注的网页内容：文本中的词、短语；图片；文档等。
- RDFa是W3C推荐的一种将RDF嵌入HTML的方式

语义标注 - RDFa

```
<html>
  <head><title>An RDFa sample</title></head>
  <body>
    <p>The author of UML Distilled is Martin Fowler</p>
  </body>
</html>
```

```
<html xmlns="http://www.w3.org/1999/xhtml" xmlns:ex="http://example.org/" version="XHTML+RDFa
1.0" xml:lang="en">
  <head><title>An RDFa sample</title></head>
  <body>
    <p> <div typeof="ex:Book" about="http://example.org/umlId">
      The author of <span property="ex:title">UML Distilled</span> is
      <div typeof="ex:Person" about="http://example.org/mfowler">
        <span property="ex:name">Martin</span> <span property="ex:surname">Fowler</span>
      </div>
    </div></p>
  </body>
</html>
```

OWL

- OWL (Ontology Web Language) 是一种用于表示本体的语言，被用于支持Semantic Web的构建
 - 在2004年成为W3C推荐
 - 在语法上，OWL本体是有效的RDF、XML文档
 - OWL基于描述逻辑
 - OWL Lite对应描述逻辑语言SHIF(D)
 - OWL DL对应描述逻辑语言SHOIN(D)
- OWL的基本元素：类（class）、属性（property）、对象（individual）

OWL – 例

```
<!DOCTYPE rdf:RDF [<!ENTITY owl "http://www.w3.org/2002/07/owl#">]>
<rdf:RDF xmlns:owl ="http://www.w3.org/2002/07/owl#"
  xmlns:rdf ="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
  <owl:Class rdf:ID="Person" />
  <owl:Class rdf:ID="Man">
    <rdfs:subClassOf rdf:resource="#Person" />
  </owl:Class>
  <owl:ObjectProperty rdf:ID="hasChild" />
  <owl:ObjectProperty rdf:ID="hasDaughter">
    <rdfs:subPropertyOf rdf:resource="#hasChild" />
  </owl:ObjectProperty>
  <owl:DatatypeProperty rdf:ID="age" />
</rdf:RDF>
```


OWL的基本元素

- 类 (Classes)
 - 表示类别、概念，是个体的集合
 - 如：人、地方、男人等
- OWL可表示类的子类关系 (subclass)、等价 (equivalent)、不相交 (disjoint)、交 (intersection)、并 (union) 等

```
SubClassOf( :Mother :Woman )  
EquivalentClasses( :Person :Human )  
DisjointClasses( :Woman :Man )
```

```
EquivalentClasses(  
  :Mother  
  ObjectIntersectionOf( :Woman :Parent )  
)
```

OWL的基本元素

- 属性 (Property)
 - 对象属性 (object property)
 - 描述两个对象间的关系
 - 数据类型属性 (datatype property)
 - 将对象与具体值 (literals, 如字符串、数值等) 关联
- OWL可描述属性的层级关系、对属性的限制

```
SubObjectPropertyOf( :hasWife :hasSpouse )  
ObjectPropertyDomain( :hasWife :Man )  
ObjectPropertyRange( :hasWife :Woman )
```

OWL的基本元素

- 对象 (individual)
- OWL可表示对象属于某类、对象间的关系、对象的属性等

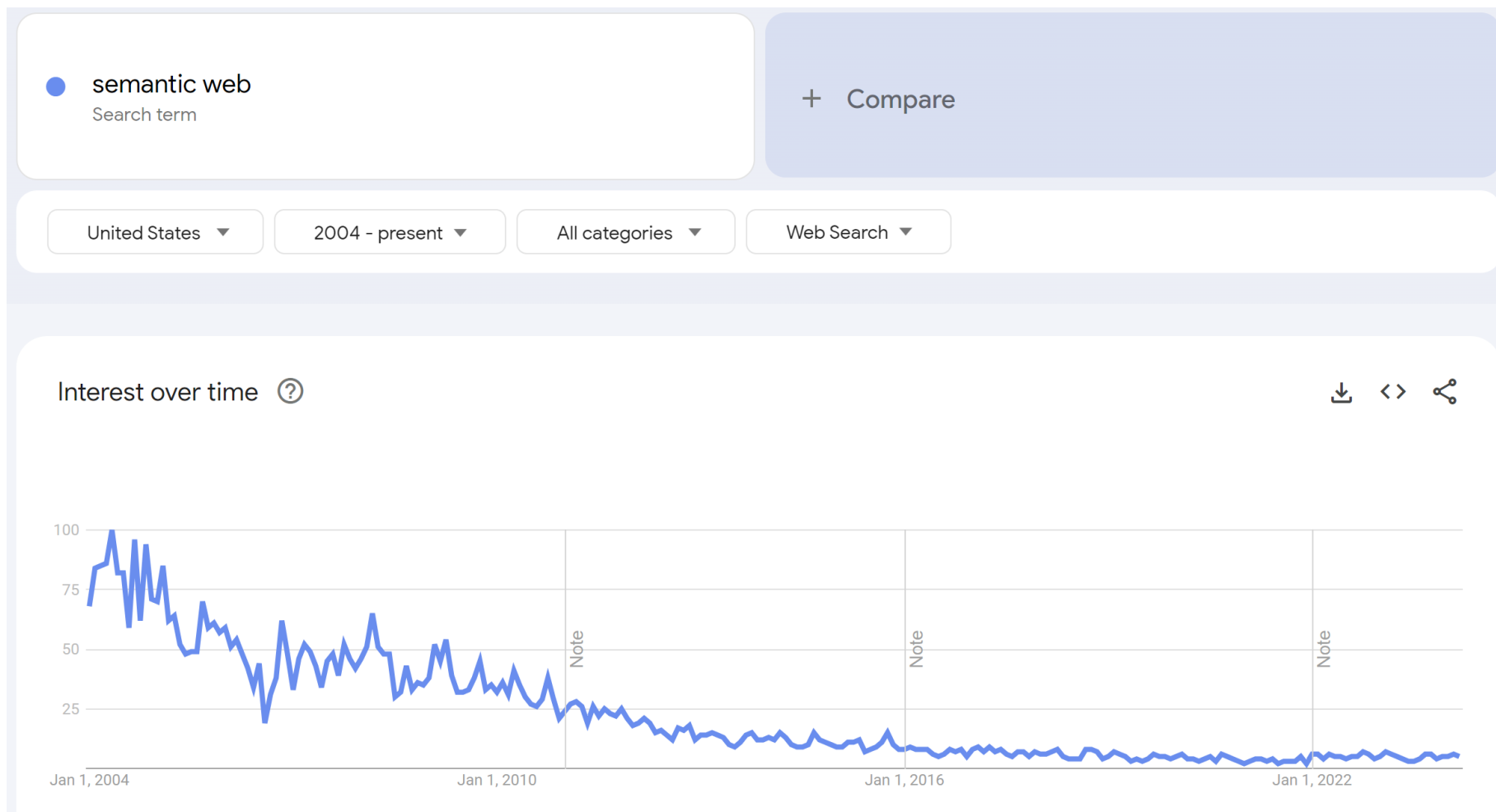
```
ClassAssertion( :Person :Mary )  
DifferentIndividuals( :John :Bill )  
DataPropertyAssertion( :hasAge :John "51"^^xsd:integer )
```

关联开放数据

- 构建机器可读的，可关联的数据的前提：
 - 1) 对所有资源（人、书、电影、概念）赋予唯一的标识
 - URI
 - 2) 一个统一的对知识进行表示和解释的体系
 - XML、RDF、OWL等

语义网的窘境

- 语义网的热度已大不如前，似乎开始被大家遗忘

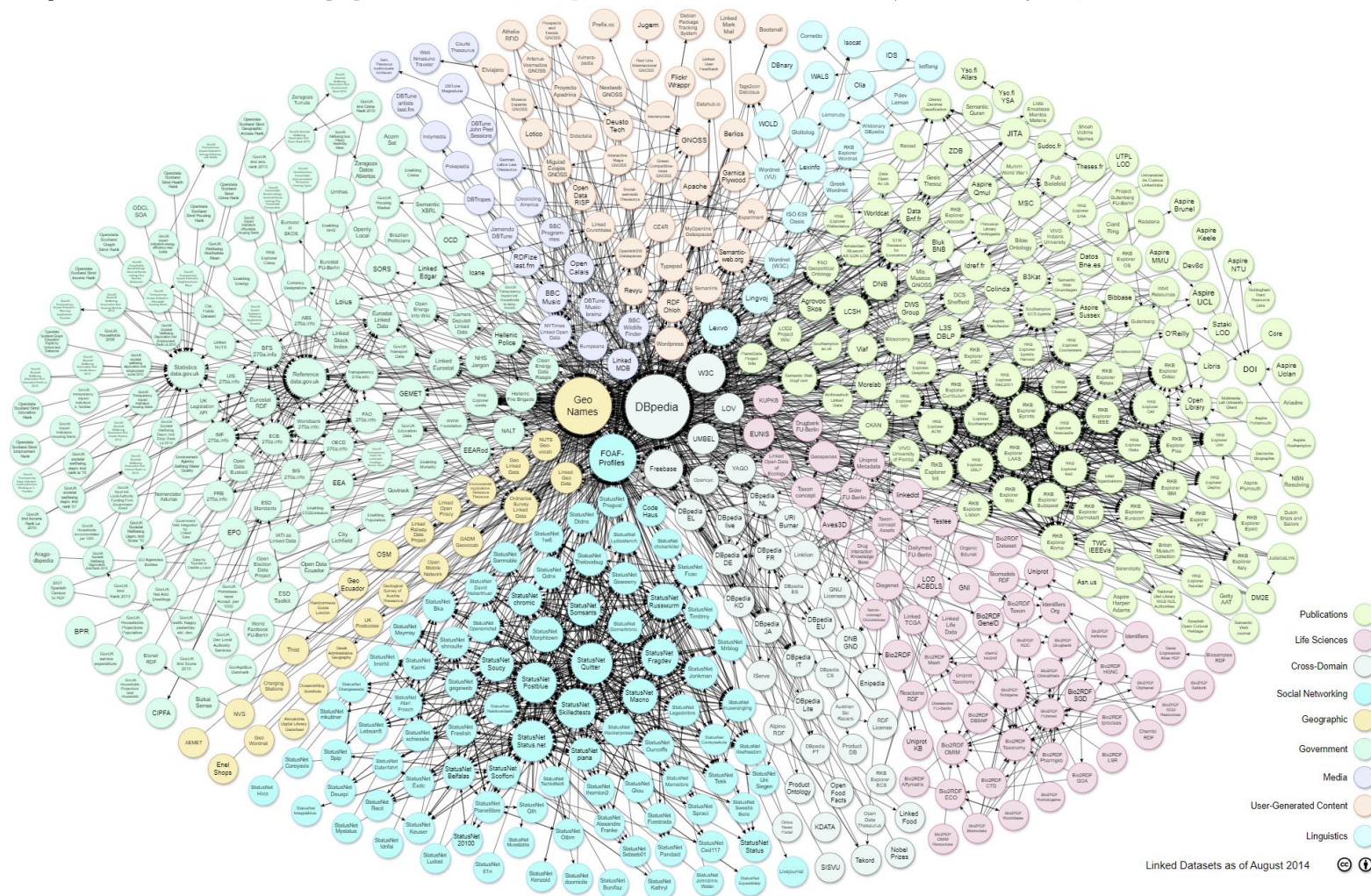


语义网的窘境

- 依赖大家自愿地提供高质量的数据和标注
 - 发布信息的成本提高但收益很有限
 - “I found it difficult to provide a reasonable business justification to my manager.”
 - 无法阻止大家提供错误的数据或标注
 - 同时机器的自然语言理解能力在不断提高
- W3C提供了很多标准和建议，但它们很复杂
 - “I studied semantic web for a semester during college and I found it was too complicated and confusing”

语义网的窘境

- 虽然如此，它已经给我们提供了一些有用的资源



END
