Semantic Technologies: Representing Semantic Data

Dr Matthew Rowe
Research Associate, Knowledge Media Institute, The Open University, UK

http://people.kmi.open.ac.uk/rowe
http://www.matthew-rowe.com
@mattroweshow

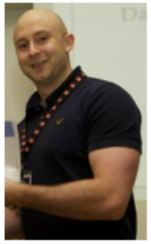
Outline

- Introduction
  - Why do we need semantics?
- Semantic Technologies
  - Representing Knowledge
    - Metadata Models
    - Ontologies
  - Querying and Inference
- Semantics in the Wild
  - Ontologies/vocabularies
  - Applications of Semantic Technologies

Why do we need Semantics?

The World Wide Web

- A flat document network
- Documents are associated via hyperlinks
  - No meaning assigned to links
    - We know that page A refers to page B
  - Documents refer to entities within
    - However the WWW ignores this, or rather:
    - This is not explicit to machines!
- Humans can read information in web pages…
  - But machines find this hard to do
    - Lots of noise
    - Complexity of semi-structured documents



Dr Matthew Rowe

Email: m.c.rowe@open.ac.uk
Phone: +44 (0)1908 653169
Address: Knowledge Media Institute, The Open University, Milton Keynes, UK, MK7 6AA
URI: http://people.kmi.open.ac.uk/rowe/foaf.rdf#me

2.0 arctic armandalannucci better brooker bubble cameras charlie chrismorris circular cool cover data evolution facebook fraid, friend greets identity journal joy lcd lessevil library like microformats monkeys myspace networks oauth openid opensocial search semanticweb social socialnetworks sociology software software/web2.0/video soundsystem studies syndrome tags talking temptation user web web2.0 world you

My research interests centre around the Social Semantic Web and how information can be processed and interpreted by machines via formal semantics. In particular my work focuses on automated disambiguation methods using machine learning techniques, and information extraction and the enrichment of data with semantics.

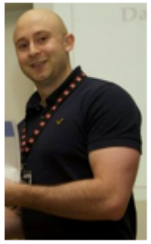
I work as a Research Associate at KMi on the EU funded project ROBUST. On this project I am the work packge leader for WP3, the work within which will create models for user and community behaviour. Through these models users will be categorised based on their behaviour. Research will be conducted into detecting trends in behaviour of both communities and individuals, allowing the evolution of communities to be predicted based on certain distinct features.

Before joining KMi I studied for a Ph.D. under the supervision of Prof. Fabio Ciravegna at the University of Sheffield. My thesis explored automated identity disambiguation methods and how such methods could be supported using Social Data leveraged from Social Web platforms such as Facebook, MySpace and Twitter. Prior to my Ph.D. studies I studied for a Masters in Software Engineering at the University of Sheffield.

Recent Publications

- Aligning Tweets with Events: Automation via Semantics. M Rowe and M Stankovic. To appear in the Semantic Web Journal. (2011)
- Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages. M Rowe, M Stankovic, A Dadzie and M Hardey. Extended Semantic Web Conference 2011. Heraklion, Crete. (2011)
- Approaches to Visualising Linked Data: A Survey. A Dadzie and M Rowe. To appear in the Semantic Web Journal. (2011)
- The Effect of User Features on Churn in Social Networks (**Winner of Best Paper**). M Karnstedt, M Rowe, J Chan, H Alani and C Hayes. In proceedings of Web Science Conference 2011. Koblenz, Germany. (2011)
- Predicting Discussions on the Social Semantic Web. M Rowe, S Angeletou and H Alani. In proceedings of Extended Semantic Web Conference 2011. Heraklion, Crete. (2011)

For a full list of my publications click here.



Dr Matthew Rowe

Email: m.c.rowe@open.ac.uk
Phone: +44 (0)1908 653169
Address: Knowledge Media Institute, The Open University, Milton Keynes, UK, MK7 6AA
URI: http://people.kmi.open.ac.uk/rowe/foaf.rdf#me

Information about an entity (Person)

2.0 arctic armandalannucci better brooker bubble cameras charlie chrismorris circular cool cover data evolution facebook fraid, friend greets identity journal joy lcd lessevil library like microformats monkeys myspace networks oauth openid opensocial search semanticweb social socialnetworks sociology software software/web2.0/video soundsystem studies syndrome tags talking temptation user web web2.0 world you

My research interests centre around the Social Semantic Web and how information can be processed and interpreted by machines via formal semantics. In particular my work focuses on automated disambiguation methods using machine learning techniques, and information extraction and the enrichment of data with semantics.

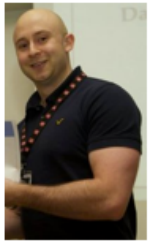
I work as a Research Associate at KMi on the EU funded project ROBUST. On this project I am the work packge leader for WP3, the work within which will create models for user and community behaviour. Through these models users will be categorised based on their behaviour. Research will be conducted into detecting trends in behaviour of both communities and individuals, allowing the evolution of communities to be predicted based on certain distinct features.

Before joining KMi I studied for a Ph.D. under the supervision of Prof. Fabio Ciravegna at the University of Sheffield. My thesis explored automated identity disambiguation methods and how such methods could be supported using Social Data leveraged from Social Web platforms such as Facebook, MySpace and Twitter. Prior to my Ph.D. studies I studied for a Masters in Software Engineering at the University of Sheffield.

Recent Publications

- Aligning Tweets with Events: Automation via Semantics. M Rowe and M Stankovic. To appear in the Semantic Web Journal. (2011)
- Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages. M Rowe, M Stankovic, A Dadzie and M Hardey. Extended Semantic Web Conference 2011. Heraklion, Crete. (2011)
- Approaches to Visualising Linked Data: A Survey. A Dadzie and M Rowe. To appear in the Semantic Web Journal. (2011)
- The Effect of User Features on Churn in Social Networks (**Winner of Best Paper**). M Karnstedt, M Rowe, J Chan, H Alani and C Hayes. In proceedings of Web Science Conference 2011. Koblenz, Germany. (2011)
- Predicting Discussions on the Social Semantic Web. M Rowe, S Angeletou and H Alani. In proceedings of Extended Semantic Web Conference 2011. Heraklion, Crete. (2011)

For a full list of my publications click here.



Dr Matthew Rowe

Name

Email: m.c.rowe@open.ac.uk
Phone: +44 (0)1908 653169
Address: Knowledge Media Institute, The Open University, Milton Keynes, UK, MK7 6AA
URI: http://people.kmi.open.ac.uk/rowe/foaf.rdf#me

2.0 arctic armandalannucci better brooker bubble cameras charlie chrismorris circular cool cover data evolution facebook fraid, friend greets identity journal joy lcd lessevil library like microformats monkeys myspace networks oauth openid opensocial search semanticweb social socialnetworks sociology software software/web2.0/video soundsystem studies syndrome tags talking temptation user web web2.0 world you

My research interests centre around the Social Semantic Web and how information can be processed and interpreted by machines via formal semantics. In particular my work focuses on automated disambiguation methods using machine learning techniques, and information extraction and the enrichment of data with semantics.

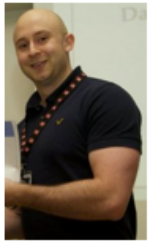
I work as a Research Associate at KMi on the EU funded project ROBUST. On this project I am the work packge leader for WP3, the work within which will create models for user and community behaviour. Through these models users will be categorised based on their behaviour. Research will be conducted into detecting trends in behaviour of both communities and individuals, allowing the evolution of communities to be predicted based on certain distinct features.

Before joining KMi I studied for a Ph.D. under the supervision of Prof. Fabio Ciravegna at the University of Sheffield. My thesis explored automated identity disambiguation methods and how such methods could be supported using Social Data leveraged from Social Web platforms such as Facebook, MySpace and Twitter. Prior to my Ph.D. studies I studied for a Masters in Software Engineering at the University of Sheffield.

Recent Publications

- Aligning Tweets with Events: Automation via Semantics. M Rowe and M Stankovic. To appear in the Semantic Web Journal. (2011)
- Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages. M Rowe, M Stankovic, A Dadzie and M Hardey. Extended Semantic Web Conference 2011. Heraklion, Crete. (2011)
- Approaches to Visualising Linked Data: A Survey. A Dadzie and M Rowe. To appear in the Semantic Web Journal. (2011)
- The Effect of User Features on Churn in Social Networks (**Winner of Best Paper**). M Karnstedt, M Rowe, J Chan, H Alani and C Hayes. In proceedings of Web Science Conference 2011. Koblenz, Germany. (2011)
- Predicting Discussions on the Social Semantic Web. M Rowe, S Angeletou and H Alani. In proceedings of Extended Semantic Web Conference 2011. Heraklion, Crete. (2011)

For a full list of my publications click here.



Dr Matthew Rowe

Name

Email: m.c.rowe@open.ac.uk

Email

Phone: +44 (0)1908 653169

Address: Knowledge Media Institute, The Open University, Milton Keynes, UK, MK7 6AA

URI: http://people.kmi.open.ac.uk/rowe/foaf.rdf#me

2.0 arctic armandalannucci better brooker bubble cameras charlie chrismorris circular cool cover data evolution facebook fraid, friend greets identity journal joy lcd lessevil library like microformats monkeys myspace networks oauth openid opensocial search semanticweb social socialnetworks sociology software software/web2.0/video soundsystem studies syndrome tags talking temptation user web web2.0 world you

My research interests centre around the Social Semantic Web and how information can be processed and interpreted by machines via formal semantics. In particular my work focuses on automated disambiguation methods using machine learning techniques, and information extraction and the enrichment of data with semantics.

I work as a Research Associate at KMi on the EU funded project ROBUST. On this project I am the work packge leader for WP3, the work within which will create models for user and community behaviour. Through these models users will be categorised based on their behaviour. Research will be conducted into detecting trends in behaviour of both communities and individuals, allowing the evolution of communities to be predicted based on certain distinct features.

Before joining KMi I studied for a Ph.D. under the supervision of Prof. Fabio Ciravegna at the University of Sheffield. My thesis explored automated identity disambiguation methods and how such methods could be supported using Social Data leveraged from Social Web platforms such as Facebook, MySpace and Twitter. Prior to my Ph.D. studies I studied for a Masters in Software Engineering at the University of Sheffield.

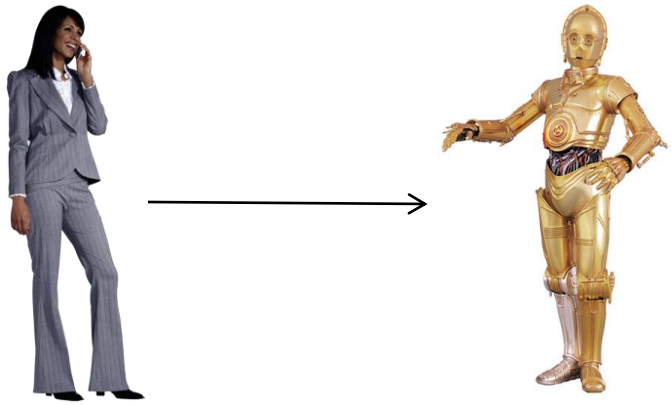
Recent Publications

- Aligning Tweets with Events: Automation via Semantics. M Rowe and M Stankovic. To appear in the Semantic Web Journal. (2011)
- Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages. M Rowe, M Stankovic, A Dadzie and M Hardey. Extended Semantic Web Conference 2011. Heraklion, Crete. (2011)
- Approaches to Visualising Linked Data: A Survey. A Dadzie and M Rowe. To appear in the Semantic Web Journal. (2011)
- The Effect of User Features on Churn in Social Networks (**Winner of Best Paper**). M Karnstedt, M Rowe, J Chan, H Alani and C Hayes. In proceedings of Web Science Conference 2011. Koblenz, Germany. (2011)
- Predicting Discussions on the Social Semantic Web. M Rowe, S Angeletou and H Alani. In proceedings of Extended Semantic Web Conference 2011. Heraklion, Crete. (2011)

For a full list of my publications click here.

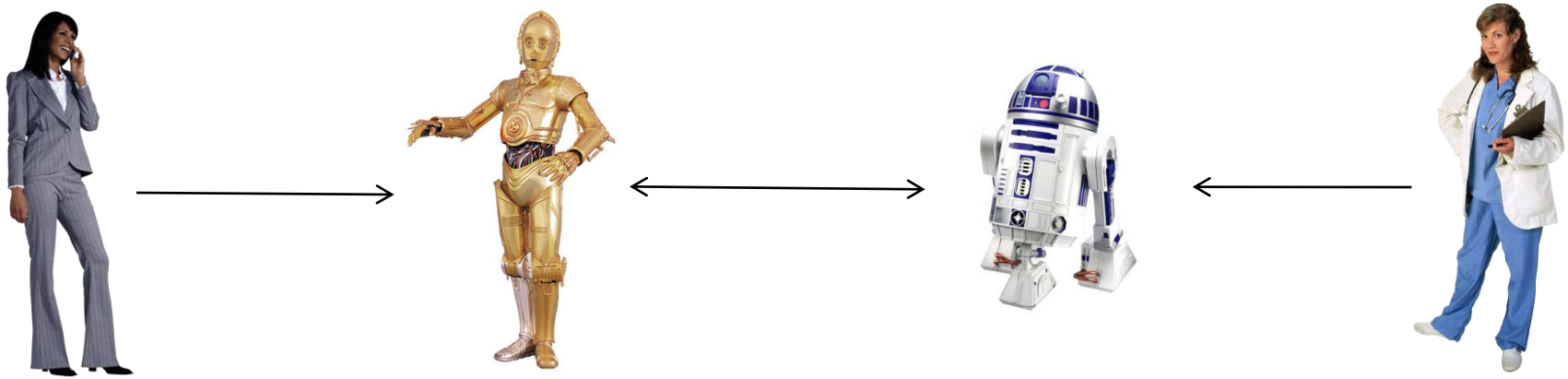
Initial Semantic Web Vision

From [Berners-Lee et al, 2001]



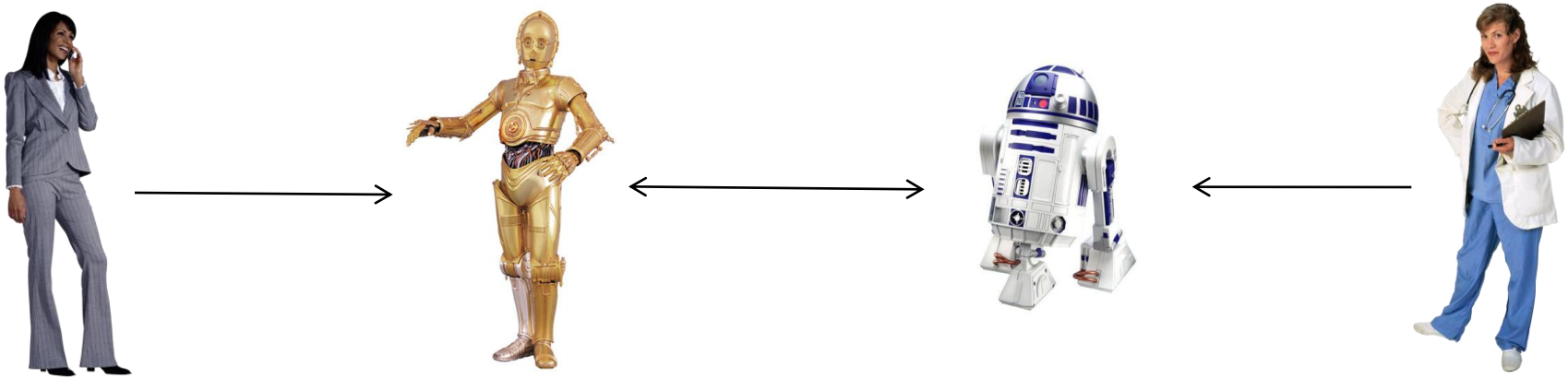
- Alice's mom (sic) needs to book a physiotherapy appointment
- She uses her software agent to handle the booking

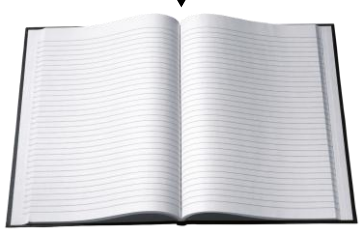
Initial Semantic Web Vision



- The agent liaises with the physiotherapist's software agent …

Initial Semantic Web Vision



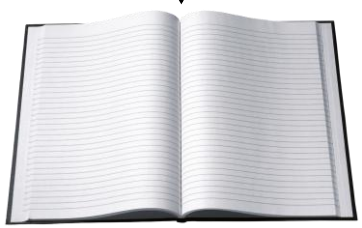


…and Alice's diary

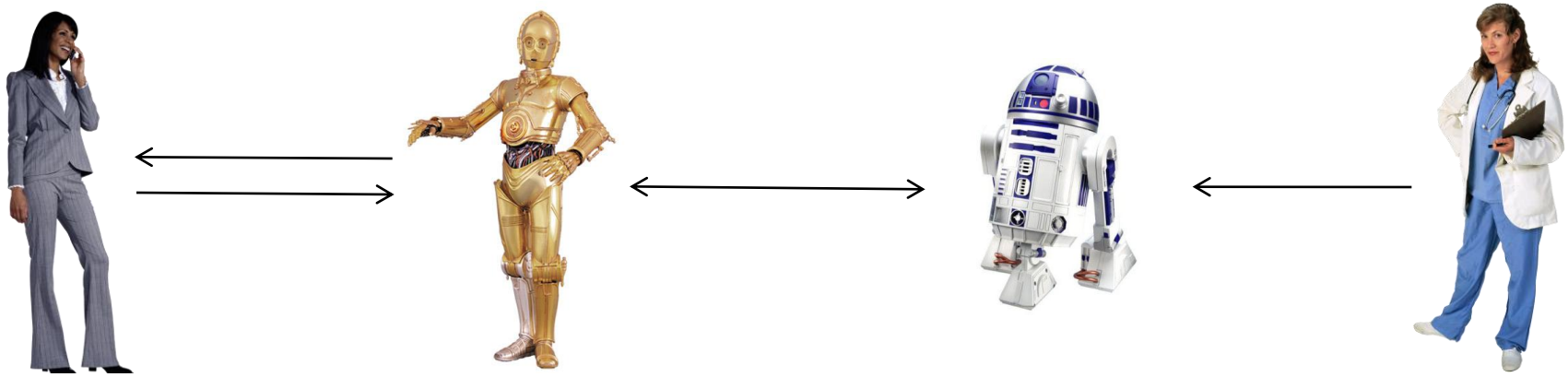
Initial Semantic Web Vision



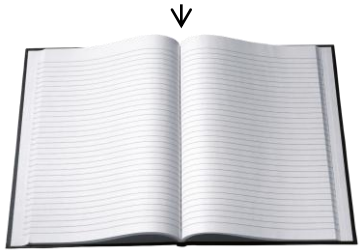

- Also liaises with her brother's software agent

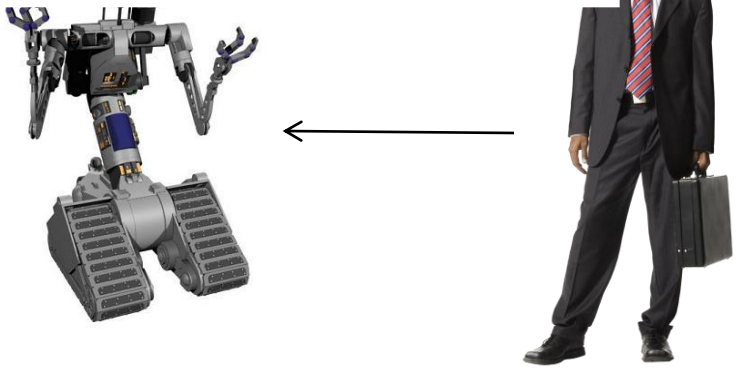


Initial Semantic Web Vision



- Returns a list of suitable appointments
- Alice chooses the one that suits her





The Semantic Web

- The WWW does not support this initial vision
- Information is combined together in a single web page
  - No distinction between entities within the page
- Semantic Web is founded on three key ideas:
  - Express meaning
  - Ontologies
  - Agents
- Documents contain entities or 'things'
- Ergo… define the semantics of the entity and link entities together
  - Define the semantics of the link
    - i.e. its meaning, what does the association tell us?

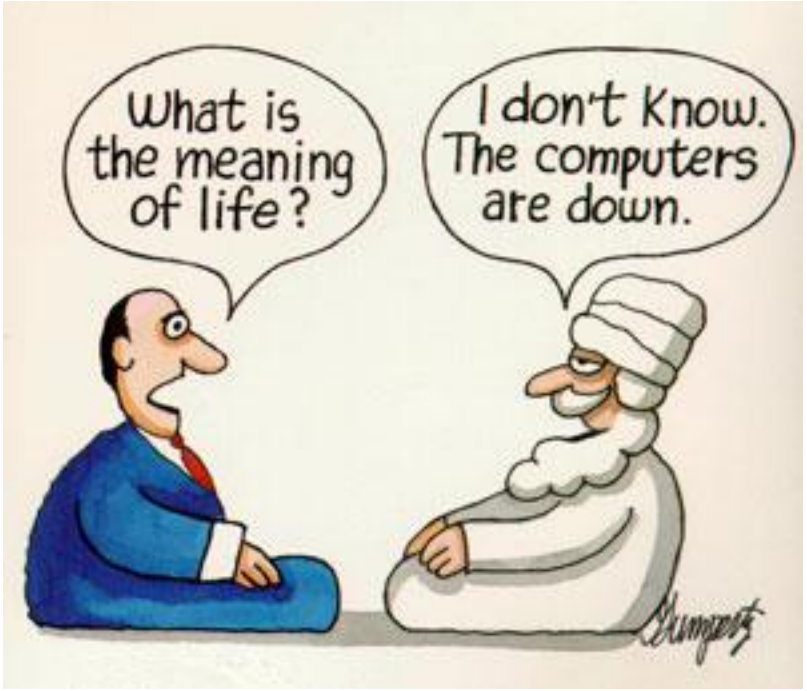
A Modern Semantic Web Use Case

- Alice is a keen online shopper, and so are her friends
- She wants product suggestions based on what her friends are talking about
- Her software agent gathers her profile information from several Social Web platforms
- The agent analyses what her friends are talking about and like
- Liked products are found on several shopping sites
- Alice is shown the selection of products, ranked by price, and **why** they were selected

We will revisit this use case later

Enable Understanding

- Machines are like humans, they transfer information
- However… information is provided in different formats
  - i.e. different languages
- How can machines understand information that differs from its known format?
  - E.g. data from one API differs from another
- N.b. protocol droids haven't been invented…. yet



The provision of meaning

Data, Information, Knowledge and Wisdom

- Data: raw characters and symbols
- Information: data in a usable form
  - Can answer: who, what, where?
    - E.g. who is Alice? Where does Alice live?
- Knowledge: information enriched with semantics
  - Can answer: how?
    - E.g. how does Alice know Bob?
- Wisdom: evaluated understanding, hindsight

Representing Knowledge

- Metadata is required that can be understood by machines
  - Metadata = data about data
- Metadata descriptions must be commonly interpretable
  - E.g. machine A knows what machine B is talking about
- Data can differ, but metadata should described in a consistent form
  - Heterogeneity of data, homogeneity of metadata

Metadata Models: XML

- What is it?
  - eXtensible Markup Language
  - Enables the definition of metadata models using user-defined markup language
  - Developers define bespoke tags for bespoke purposes
  - Information is structured in a nested format
- Limitations:
  - Vocabulary of tags must be known by both parties for information transfer
  - Meaning of tags is obvious to humans but not to machines
  - Restricted to a hierarchical structure

XML Example: Twitter API Response

```
<status>
  <created_at>Mon Jul 18 14:50:12 +0000 2011</created_at>
  <id>92696422590976000</id>
  <text>Skills are obtained by that constant, sickening hours of practice on your talen
  <source>&lt;a href="http://twitter.com/#!/download/ipad" rel="nofollow"&gt;Twitter fo
  <truncated>false</truncated>
  <favorited>false</favorited>
  <in_reply_to_status_id></in_reply_to_status_id>
  <in_reply_to_user_id></in_reply_to_user_id>
  <in_reply_to_screen_name></in_reply_to_screen_name>
  <retweet_count>0</retweet_count>
  <retweeted>false</retweeted>
  <user>
    <id>26741696</id>
    <name>Frank Puriza</name>
    <screen_name>turaboyy</screen_name>
    <location>San Diego</location>
    <description>Pastor... Priest... Athlete... Bishop... Pimp... Lawyer... Son... Doct
Mechanic...</description>
    <profile_image_url>http://a0.twimg.com/profile_images/1435612218/IMG_0104_normal.jp
    <profile_image_url_https>https://si0.twimg.com/profile_images/1435612218/IMG_0104_n
    <url></url>
    <protected>false</protected>
    <followers_count>342</followers_count>
```

What is a user?

Metadata Models: XML Schema & Namespaces

- What is it?
  - XML Schema defines the structure that an XML document must adhere to
  - Namespaces are used to define the location of the schema
  - Prefixes are used to abbreviate the namespace
    - E.g. xmlns:xs="http://www.w3.org/2001/XMLSchema"
      - Prefix = xs, Namespace = http://www.w3.org/2001/XMLSchema
  - Uniform Resource Identifiers are used to refer to elements in the schema
    - xs:elementName = http://www.w3.org/2001/XMLSchema#elementName
- Limitations:
  - Still restricts XML to a tree-like structure
  - Semantics are implicit to the schema
  - Cannot extend and combine schemas

XML Schema Example

```
<author>
  <name>GoogleDevelopers</name>
  <uri>https://gdata.youtube.com/feeds/api/users/GoogleDevelopers</uri>
</author>
<yt:firstName>33</yt:firstName>
<yt:lastName>Jones</yt:lastName>
<yt:aboutMe>My favorite number is 33.</yt:aboutMe>
<yt:age>33</yt:age>
<yt:username>GoogleDevelopers</yt:username>
<yt:books>Catch-22</yt:books>
<yt:gender>m</yt:gender>
<yt:company>Google</yt:company>
<yt:hobbies>Testing YouTube APIs</yt:hobbies>
<yt:hometown>Philadelphia, PA</yt:hometown>
<yt:location>Boston, MA 02043, US</yt:location>
<yt:movies>Aqua Teen Hungerforce</yt:movies>
<yt:music>Elliott Smith</yt:music>
<yt:relationship>taken</yt:relationship>
<yt:occupation>Technical Writer</yt:occupation>
<yt:school>University of North Carolina</yt:school>
```

xmlns:yt="http://gdata.youtube.com/schemas/2007"

Metadata Models: RDF

- What is it?
  - Resource Description Format
  - Graph-based metadata model representation
  - Consists of three main concepts:
    - Resources
    - Properties
    - Statements
  - Resources:
    - An object or thing identified by a URI
      - E.g. A person can be defined by:
        <http://people.kmi.open.ac.uk/rowe/foaf.rdf#me>
  - Properties:
    - Describe attributes or relations of a resource
      - E.g. has_name, has_email

Metadata Models: RDF

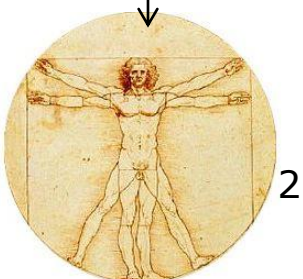
- Statements:
  - Consists of a resource, a property and a value of that property
  - AKA triples: subject, predicate, object
  - E.g. Matthew Rowe is a person

<http://people.kmi.open.ac.uk/rowe/foaf.rdf#me>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://xmlns.com/foaf/0.1/Person>

<http://people.kmi.open.ac.uk/rowe/foaf.rdf#me>

<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>

<http://xmlns.com/foaf/0.1/Person>



Metadata Models: RDF

- Statements can be linked together
  - Object of one triple becomes the subject of another triple

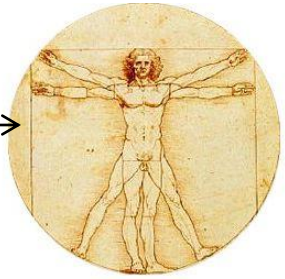
<http://people.kmi.open.ac.uk/rowe/foaf.rdf#me>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://xmlns.com/foaf/0.1/Person> ;
<http://people.kmi.open.ac.uk/rowe/foaf.rdf#me>
<http://xmlns.com/foaf/0.1/knows>
<http://people.kmi.open.ac.uk/rowe/foaf.rdf#591280907>

Metadata Models: RDF

<http://people.kmi.open.ac.uk/rowe/foaf.rdf#me>

<http://xmlns.com/foaf/0.1/Person>

<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>



<http://xmlns.com/foaf/0.1/knows>

<http://people.kmi.open.ac.uk/rowe/foaf.rdf#591280907>

Metadata Models: RDF

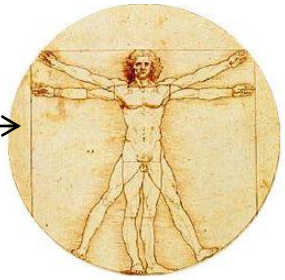
<http://people.kmi.open.ac.uk/rowe/foaf.rdf#me>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://xmlns.com/foaf/0.1/Person> ;
<http://people.kmi.open.ac.uk/rowe/foaf.rdf#me>
<http://xmlns.com/foaf/0.1/knows>
<http://people.kmi.open.ac.uk/rowe/foaf.rdf#591280907> ;
<http://people.kmi.open.ac.uk/rowe/foaf.rdf#591280907>
<http://xmlns.com/foaf/0.1/based_near>
<http://sws.geonames.org/2643743> ;
<http://sws.geonames.org/2643743>
<http://www.geonames.org/ontology#name>
"London"

Metadata Models: RDF

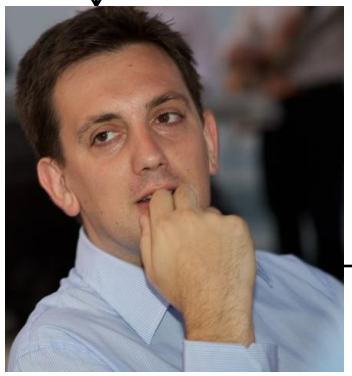
<http://people.kmi.open.ac.uk/rowe/foaf.rdf#me>

<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>

<http://xmlns.com/foaf/0.1/Person>



<http://xmlns.com/foaf/0.1/knows>



<http://xmlns.com/foaf/0.1/based_near>

<http://people.kmi.open.ac.uk/rowe/foaf.rdf#591280907>

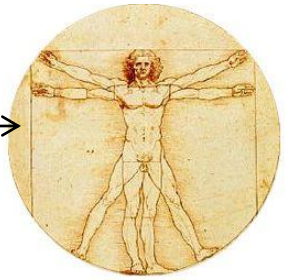
<http://sws.geonames.org/2643743>

Metadata Models: RDF

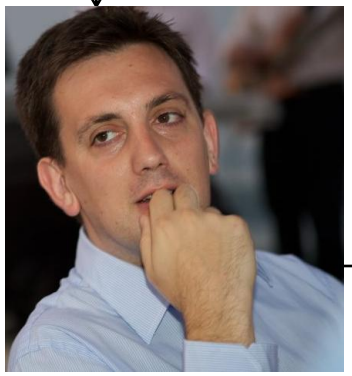
<http://people.kmi.open.ac.uk/rowe/foaf.rdf#me>

<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>

<http://xmlns.com/foaf/0.1/Person>



<http://xmlns.com/foaf/0.1/knows>



<http://xmlns.com/foaf/0.1/based_near>

<http://people.kmi.open.ac.uk/rowe/foaf.rdf#591280907>

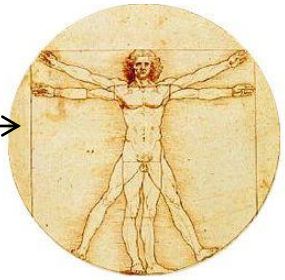
<http://sws.geonames.org/2643743>

Metadata Models: RDF

<http://people.kmi.open.ac.uk/rowe/foaf.rdf#me>

<http://xmlns.com/foaf/0.1/Person>

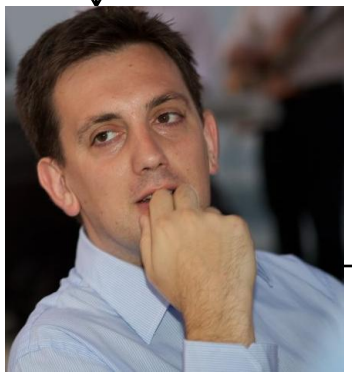
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>



"London"

<http://xmlns.com/foaf/0.1/knows>

<http://www.geonames.org/ontology#name>



<http://xmlns.com/foaf/0.1/based_near>

<http://people.kmi.open.ac.uk/rowe/foaf.rdf#591280907>

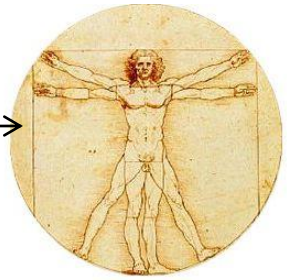
<http://sws.geonames.org/2643743>

Metadata Models: RDF

<http://people.kmi.open.ac.uk/rowe/foaf.rdf#me>

<http://xmlns.com/foaf/0.1/Person>

<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>

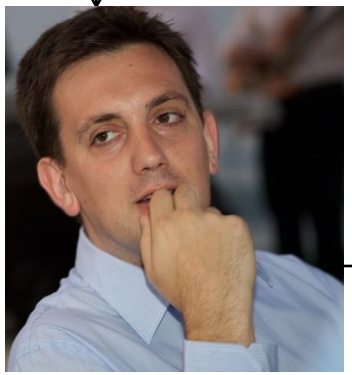


The 'Graph' is the power and utility of RDF

<http://xmlns.com/foaf/0.1/knows>

"London"

<http://www.geonames.org/ontology#name>



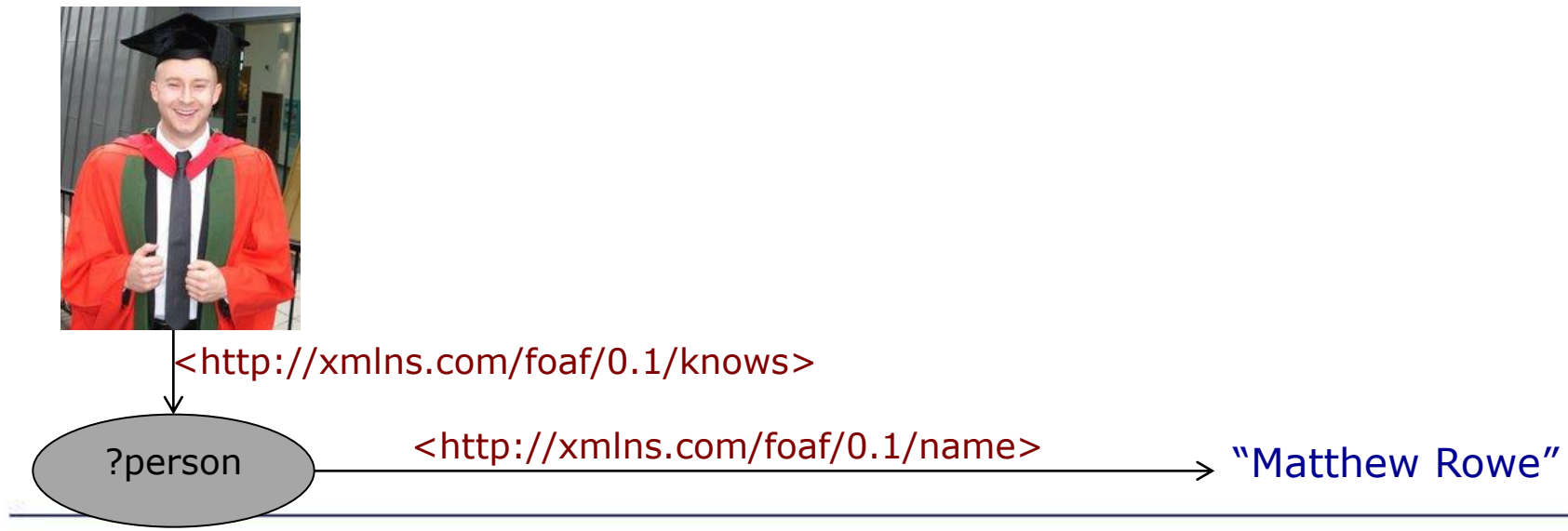
<http://xmlns.com/foaf/0.1/based_near>

<http://people.kmi.open.ac.uk/rowe/foaf.rdf#591280907>

<http://sws.geonames.org/2643743>

Metadata Models: RDF

- Blank Nodes
  - Resources can also be blank nodes
  - i.e. have no URI
    - E.g. ?person <http://xmlns.com/foaf/0.1/name> "Matthew Rowe"



Metadata Models: RDF

- Prefixes and Namespaces
  - As in XML, prefixes and namespaces are used to abbreviate concept URIs to a shorter form

<http://xmlns.com/foaf/0.1/knows> = foaf:knows

<http://www.w3.org/1999/02/22-rdf-syntax-ns#type> = rdf:type

Metadata Models: Serializing RDF

- RDF is a conceptual metadata model
  - Thus far we have only looked at triple patterns and graphs
- It can be serialized in various formats for transfer
- XML
  - Maintaining URIs for resources and namespaces for properties

```
<rdf:RDF
   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:foaf="http://xmlns.com/foaf/0.1/">
   <foaf:Person rdf:about="http://people.kmi.open.ac.uk/rowe/foaf.rdf#me”>
      <foaf:name>Matthew Rowe</foaf:name>
      <foaf:mbox rdf:resource=“mailto:m.c.rowe@open.ac.uk”/>
   </foaf:Person>
</rdf:RDF>
```

Metadata Models: Serializing RDF

- Notation 3
  - More readable form of RDF than XML

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
@prefix foaf: <http://xmlns.com/foaf/0.1/>
<http://people.kmi.open.ac.uk/rowe/foaf.rdf#me>
    foaf:name "Matthew Rowe";
    foaf:mbox <mailto:m.c.rowe@open.ac.uk> ;
```

- Turtle (Terse RDF Triple Language)
  - Enables embedding of statements within triple patterns

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
@prefix foaf: <http://xmlns.com/foaf/0.1/>
<http://people.kmi.open.ac.uk/rowe/foaf.rdf#me>
   foaf:name "Matthew Rowe" ;
   foaf:knows [
      foaf:name "Matthew Rowe"
   ] .
```

Metadata Models: RDFa

- Hang on, didn't the original semantic web vision say:

> *"...semantics were encoded into the Web page ... using off-the-shelf software for writing semantic web pages"*
>
> [Berners-Lee et al, 2001]

- The above serializations cannot be encoded within web pages, but one format can be…. RDFa
- RDFa stands for 'RDF in Attributes'
  - Allows RDF to be embedded within web page elements
  - Enriches the underlying page structure with semantics
- Standard HTML is often poorly formed and therefore cannot be parsed directly by machines
- eXtensible HyperText Markup Language (XHTML) enforces well-formed page structures, thus allowing parsing

Metadata Models: RDFa

Standard XHTML:

```
<div>
  <p>Matthew Rowe</p>
  <p><a
href="http://people.kmi.open.ac.uk/rowe/">Homepage</a></p>
</div>
```

XHTML with RDFa:

```
<div xmlns:foaf="http://xmlns.com/foaf/0.1/" typeof="foaf:Person">
  <p><span property="foaf:name">Matthew Rowe</span></p>
  <p><a rel="foaf:homepage"
href="http://people.kmi.open.ac.uk/rowe/">Homepage</a></p>
</div>
```

Metadata Models: RDFa

Standard XHTML:

```
<div>
  <p>Matthew Rowe</p>
  <p><a
href="http://people.kmi.open.ac.uk/rowe/">Homepage</a></p>
</div>
```

Outer Div element is the foaf:Person blank node (i.e. no URI)

XHTML with RDFa:

```
<div xmlns:foaf="http://xmlns.com/foaf/0.1/" typeof="foaf:Person">
  <p><span property="foaf:name">Matthew Rowe</span></p>
  <p><a rel="foaf:homepage"
href="http://people.kmi.open.ac.uk/rowe/">Homepage</a></p>
</div>
```

Metadata Models: RDFa

Standard XHTML:

```
<div>
  <p>Matthew Rowe</p>
  <p><a
href="http://people.kmi.open.ac.uk/rowe/">Homepage</a></p>
</div>
```

Outer Div element is the foaf:Person blank node (i.e. no URI)

It is typed explicitly

XHTML with RDFa:

```
<div xmlns:foaf="http://xmlns.com/foaf/0.1/" typeof="foaf:Person">
  <p><span property="foaf:name">Matthew Rowe</span></p>
  <p><a rel="foaf:homepage"
href="http://people.kmi.open.ac.uk/rowe/">Homepage</a></p>
</div>
```

Metadata Models: RDFa

Standard XHTML:

```
<div>
  <p>Matthew Rowe</p>
  <p><a
href="http://people.kmi.open.ac.uk/rowe/">Homepage</a></p>
</div>
```

Outer Div element is the foaf:Person blank node (i.e. no URI)

XHTML with RDFa:

It is typed explicitly

```
<div xmlns:foaf="http://xmlns.com/foaf/0.1/" typeof="foaf:Person">
  <p><span property="foaf:name">Matthew Rowe</span></p>
  <p><a rel="foaf:homepage"
href="http://people.kmi.open.ac.uk/rowe/">Homepage</a></p>
</div>
```

The instance is assigned properties and property values

Metadata Models: RDFa

- Allows existing data to be annotated semantically
  - Enables top-down enrichment with semantics
    - i.e. via auto-generated web pages from template engines
  - Newer Content Management Systems include RDFa automatically
- Experiencing widespread uptake
  - Examples later show the extent of this
    - Enhanced product descriptions, search and user profiling

Ontologies

- Knowledge is represented in metadata models
  - Metadata attributes meaning to data
- Metadata descriptions are provided using ontologies
- An ontology is a formal conceptualisation of some domain of knowledge
  - It is neither right nor wrong, simply a view
- Enables the sharing of a ‘world view’
  - i.e. Person A can understand how Person B relates things together
  - Or rather… machine A can understand the data that machine B is using

Ontologies

- Basic ontology is a vocabulary of terms (concepts) and their relation to one another
- Machines do not *understand* the concepts, they merely interprets the relations
- Formal constructs define restrictions on the relations
  - i.e. how two concepts can be related
- Schemas: define a document's structure
- Ontologies: conceptual description of a domain's theory

Example Ontology: FOAF

- Friend of a Friend (FOAF) Ontology
  - Used in previous examples
- Defines people and how they are related
  - i.e. who they know
- foaf:Person: class of person

  <http://people.kmi.open.ac.uk/rowe/foaf.rdf#me>

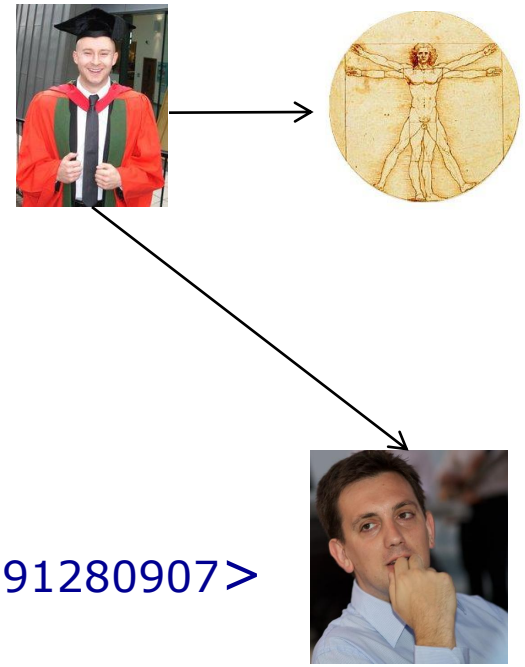
  rdf:type

  foaf:Person
- foaf:knows: property

  <http://people.kmi.open.ac.uk/rowe/foaf.rdf#me>

  foaf:knows

  <http://people.kmi.open.ac.uk/rowe/foaf.rdf#591280907>



Ontologies

- Two main ontology modelling languages are used:
  - RDF-Schema
    - Contains classes and properties
    - Property has a domain and a range
      - E.g. for foaf:name…
        - » domain: foaf:Person
        - » range: rdfs:Literal
  - Web Ontology Language (OWL)
    - Allows cardinality restrictions on properties
    - Object properties: relation between two objects or things
      - E.g. foaf:knows
    - Datatype properties: relate an object with a data value
      - E.g. foaf:name

Querying Semantic Data: SPARQL

- SPARQL Protocol and Query Language (recursive acronym!)
- Matches triple patterns within RDF
  - WHERE clause contains the patterns to match
  - SELECT clause contains variables that are to be selected
- Variables are denoted by '?'

```
PREFIX foaf:<http://xmlns.com/foaf/0.1/>
SELECT {
   ?name  }
WHERE {
   ?x foaf:name ?name
}
```

Prolog

SELECT Clause

WHERE Clause

http://www.w3.org/TR/rdf-sparql-query/

Querying Semantic Data: SPARQL

- SPARQL Protocol and Query Language (recursive acronym!)
- Matches triple patterns within RDF
  - WHERE clause contains the patterns to match
  - SELECT clause contains variables that are to be selected
- Variables are denoted by '?'

```
PREFIX foaf:<http://xmlns.com/foaf/0.1/>
SELECT {
   ?name  }
WHERE {
   ?x foaf:name ?name
}
```

Give me all person names

Querying Semantic Data: SPARQL

- URIs can also be used within queries

```
PREFIX foaf:<http://xmlns.com/foaf/0.1/>
SELECT {
        ?email, ?name }
WHERE {
  <http://people.kmi.open.ac.uk/rowe/foaf.rdf#me> foaf:knows ?person .
  ?person foaf:name ?name .
  ?person foaf:mbox ?email
}
```

Get me the names and email addresses of all people that this person knows

Inference over Semantic Data

- Based on the presence of knowledge, we can make inferences
- Rules are used to infer a conclusion given a set of premises that hold and are true

If
a person is male
and has a parent
and the parent has a sibling
then
the person is the nephew of parent's sibling

Parent(Y,X), male(X), sibling(Y,P) -> nephew(X,P)

Antecedent: premises that must hold and be true

Consequent: conclusion drawn from the antecedent

Inference over Semantic Data

- SPARQL queries can be used to build rules
  - Change the SELECT clause to a CONSTRUCT clause

```
PREFIX foaf:<http://xmlns.com/foaf/0.1/>
PREFIX gen:<http://example.com/gen>
CONSTRUCT {
   ?person gen:nephew ?relative }
WHERE {
   ?person foaf:gender "Male" .
   ?parent gen:parent ?person .
   ?parent gen:sibling ?relative .
}
```

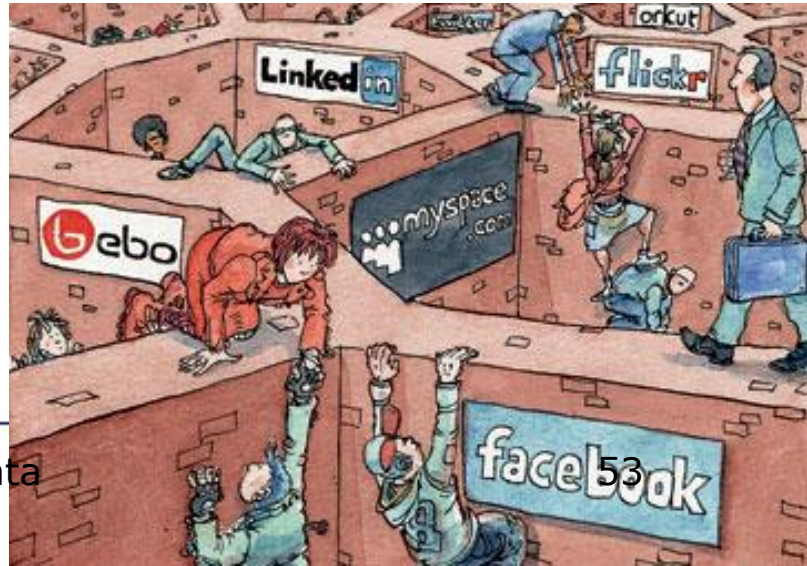
- Rules can be written in other formats
- Rules Interchange Format (RIF) transforms rules

Semantics in the Wild

Ontologies: SIOC

- Semantically Interlinked Online Communities (SIOC) Ontology
- Defines community users, forums, posts within forums and reply chains
- Defined to overcome mass-schema heterogeneity in Social Web systems
  - i.e. metadata from one platform differs from all others
- Many exporters created to leverage Social data in SIOC format

http://sioc-project.org/



Ontologies: Good Relations

- Describes commercial products and services
- Single, consolidated view of eCommerce data
  - Allows homogeneous product metadata to be unified
- Companies add Good Relations concepts to their product pages
- Used by Google for product information
  - Google crawls and extracts product information
  - Users see product information in search results

Hepp Research Personal SCSI Controller Card
★★★★★ 99 reviews - $99.99 - in stock
The Hepp Research Personal SCSI is a 16-bit add-on card that allows attaching up to seven SCSI devices to your computer. Designed in 1991 by Martin Hepp, the maker of the GoodRelations vocabulary for e-commerce.
www.heppresearch.com/commercecollator - Cached - Similar

http://purl.org/goodrelations

Vocabularies: Microformats

- Lightweight annotations embedded within XHTML structures
- hCard: represents person, company or organisation information

XHTML with RDFa:

```
<div xmlns:foaf="http://xmlns.com/foaf/0.1/" typeof="foaf:Person">
  <p><span property="foaf:name">Matthew Rowe</span></p>
  <p><a rel="foaf:homepage"
     href="http://people.kmi.open.ac.uk/rowe/">Homepage</a></p>
</div>
```

XHTML with hCard Microformat:

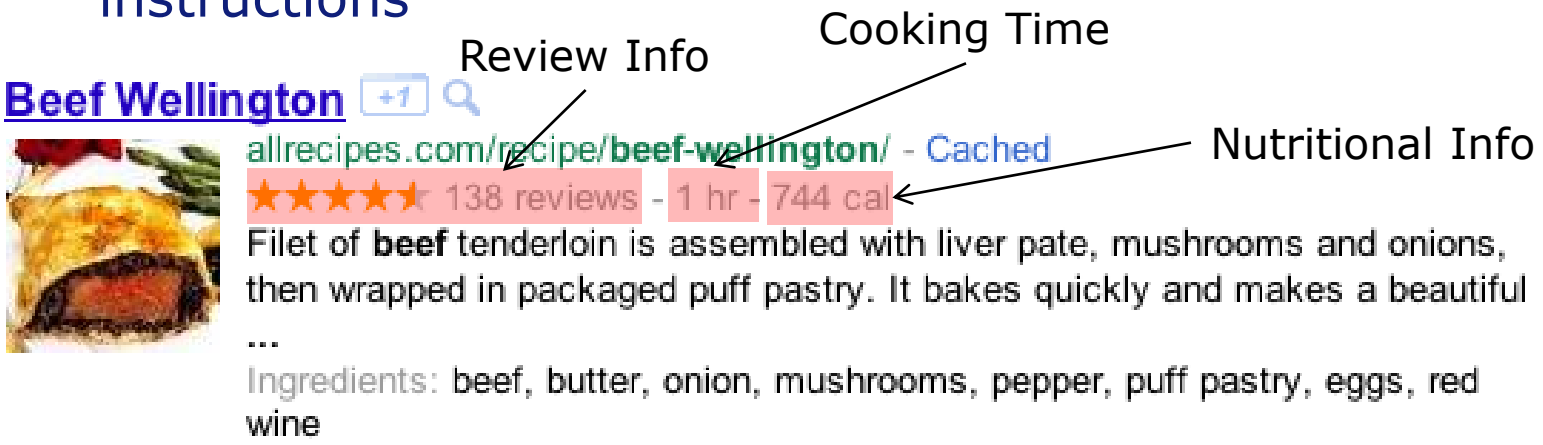
```
<div class="vcard">
  <p class="fn">Matthew Rowe</span></p>
  <p><a class="url" href="http://people.kmi.open.ac.uk/rowe/">Homepage</a></p>
</div>
```

Vocabularies: Microformats

- hCalendar: represents calendar and event information
- hReview: describes product and service reviews
- hProduct: allows embedding of product information
- XFN: semantically types hyperlinks between people
  - Specialisation of foaf:knows

```
<a href="http://people.open.ac.uk/harith" rel="colleague">Harith</a>
```

- hRecipe: describes cooking recipes (ingredients, instructions



HTML 5 & Microdata

- Microdata is a HTML 5 specification to embed semantics within web pages
  - Intended to overcome 'complexity' of RDFa & Microformats

XHTML with RDFa:

```
<div xmlns:foaf="http://xmlns.com/foaf/0.1/" typeof="foaf:Person">
  <p><span property="foaf:name">Matthew Rowe</span></p>
  <p><a rel="foaf:homepage"
     href="http://people.kmi.open.ac.uk/rowe/">Homepage</a></p>
</div>
```

Microdata:

```
<section itemscope itemtype="http://www.data-vocabulary.org/Person/">
  <p itemprop="name">Matthew Rowe</span></p>
  <p><a itemprop="url"
     href="http://people.kmi.open.ac.uk/rowe/">Homepage</a></p>
</section>
```

HTML 5 & Microdata

- Data-vocabulary.org
  - Provides definitions for Events, Organisations, Person, Product, Review
- Schema.org
  - Attempt to unify the collection of 'semantic' tags used in web pages
  - Supported by Google, Yahoo! and Bing
  - Simply switch the itemtype to use the schema's concept

RDFa

- 0.6% increase in appearance of RDFa from 2008-2009 [Mika et al, 2009]
- Facebook's Open Graph Protocol uses it!
- Drupal content management system uses it!
- Google, Yahoo! And Bing are all using it!
- So… how does it help?

Enhanced Products: Best Buy

- Motivated by users struggling to find store information and opening hours
- Solution: annotate this information within the web page
- Now uses:
  - Good Relations to describe products
  - Data-vocabulary.org to describe reviews and ratings
  - FOAF to describe people
- Increased product and service visibility
  - 30% increase in search traffic since adoption of RDFa

http://www.readwriteweb.com/archives/how_best_buy_is_using_the_semantic_web.php

Enhanced Products: Best Buy

```
<div id="content" xmlns:gr="http://purl.org/goodrelations/v1#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:foaf="http://xmlns.com/foaf/0.1/"
xmlns:v="http://rdf.data-vocabulary.org/#"
xmlns:mo="http://purl.org/ontology/mo/"
xmlns:dc="http://purl.org/dc/elements/1.1/">
...
<div rel="gr:typeOfGood">
   <div typeof="gr:ProductOrServicesSomeInstancesPlaceholder"
about="#ProductOrServicesSomeInstancesPlaceholder_9225377">
      <span property="rdfs:label" content="Apple&#174; - iPod touch&#174; 8GB* MP3 Player (4th Generation - Latest
Model) - Black">
         <span class="model" property="gr:hasMPN">MC540LL/A</span>
         <span class="sku" property="gr:hasStockKeepingUnit">9225377</span>
         <span property="gr:hasManufacturer" content="Apple&#174;"></span>
         <p class="description" property="gr:description">FaceTime camera, HD video recording, Retina display, Multi-
Touch interface; gorgeous 3.5&quot; widescreen display; Wi-Fi Web browsing</p>
         <div rel="v:hasReview">
            <div class="rating" typeof="v:Review-aggregate" about="#review_9225377">
               <span property="v:itemreviewed" content="Apple&amp;#174; - iPod touch&amp;#174; 8GB* MP3 Player
(4th Generation - Latest Model) - Black"></span>
               <span rel="v:rating">
                  <span typeof="v:Rating">
                     <span property="v:rating" datatype="xsd:string">4.7</span> of <span
property="v:best">5</span>
                     <span property="v:count" datatype="xsd:string" content="628"></span>
```

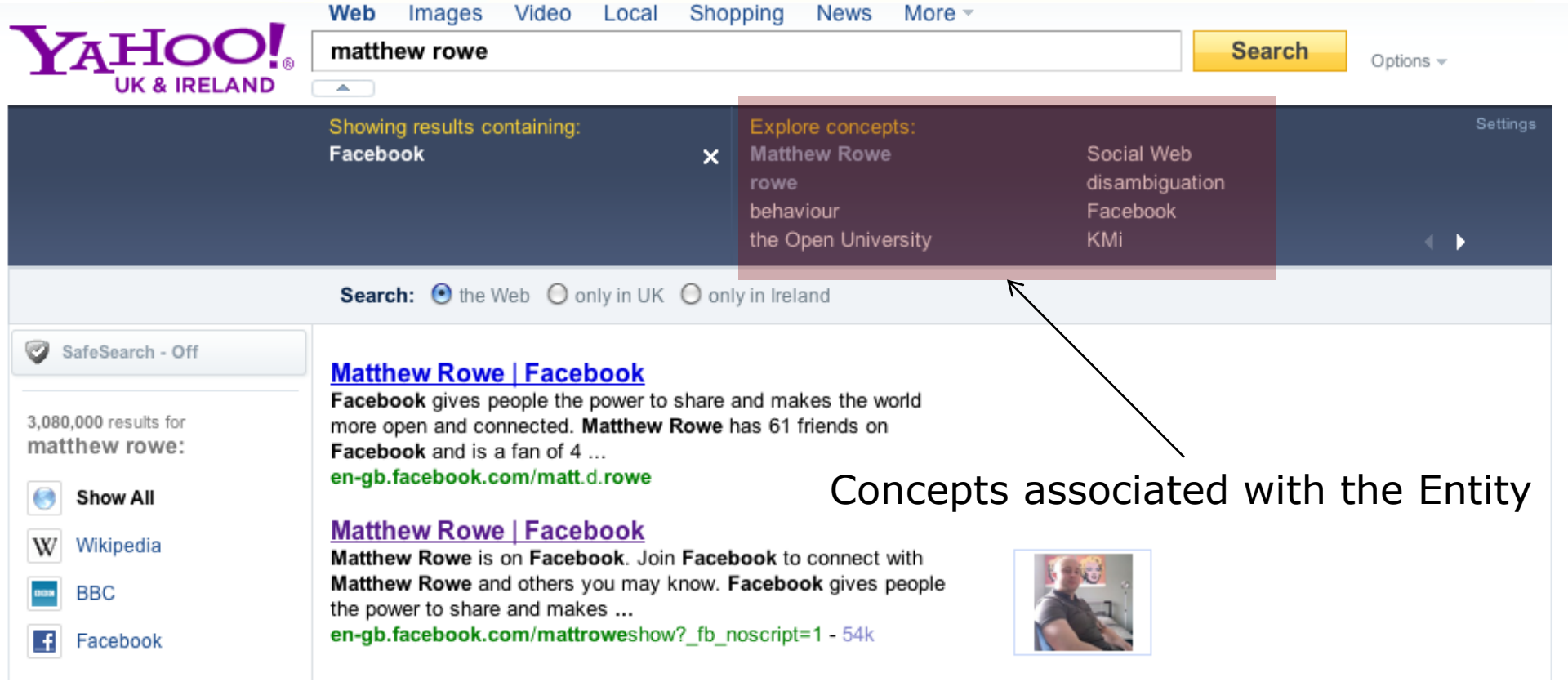
Best Buy - Apple® iPod touch® 8GB* MP3 Player (4th Generation ...
reviews.bestbuy.com › ... › All iPod & MP3 Players Reviews - Cached
★★★★½ from 630 users
Best Buy product reviews and customer ratings for Apple® **iPod touch® 8GB*** MP3 Player (4th Generation - Latest Model) - Black. Read and compare experiences ...

Enhanced Search

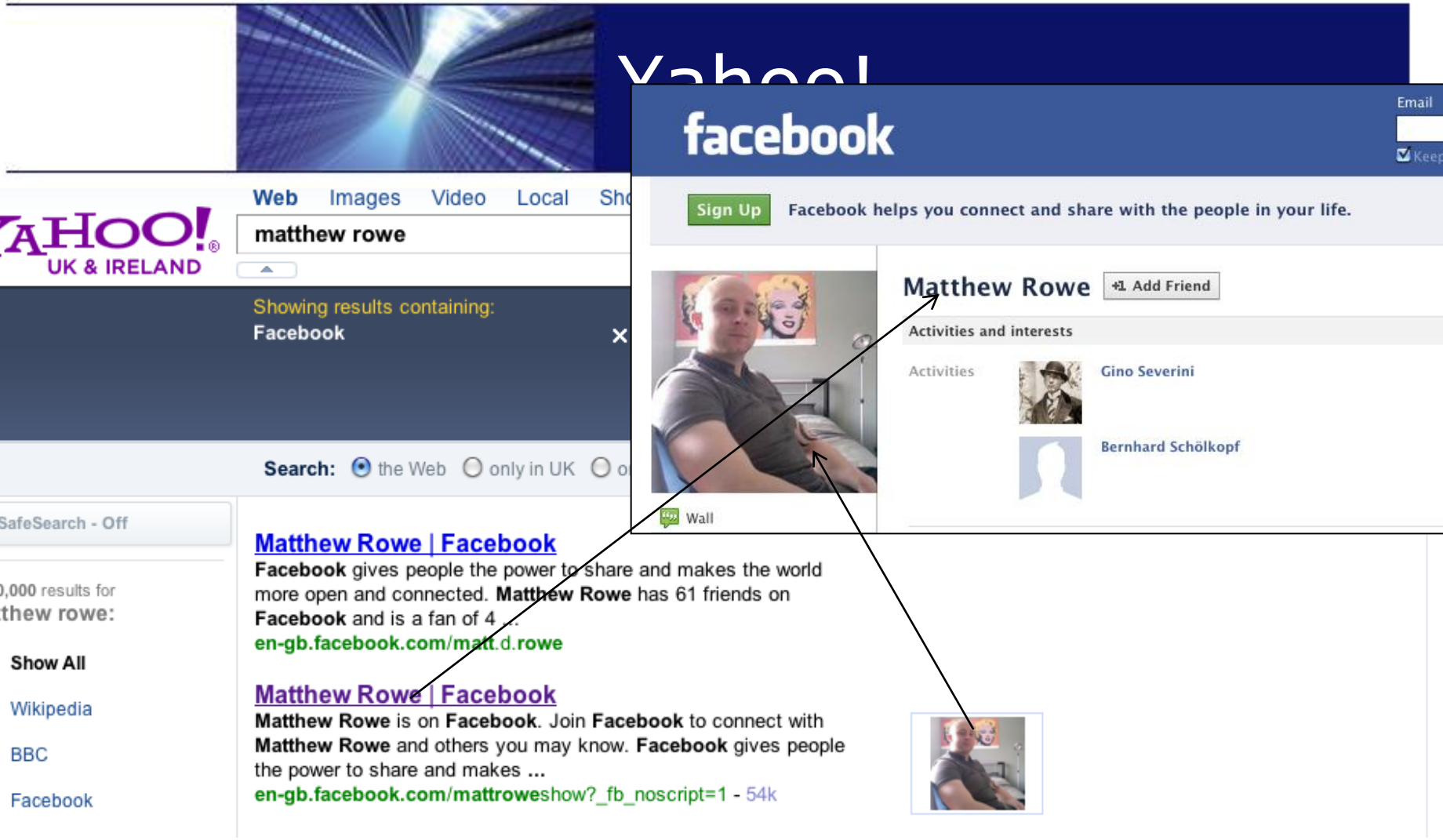
- Semantics brings information to the consumer
- No need to click further to see information
- 3 largest search engines now consume semantics to enhance search:
  - Yahoo!
  - Google Rich Snippets
  - Bing


Yahoo!
YAHOO! UK & IRELAND
Web Images Video Local Shopping News More
matthew rowe
Search
Options
Showing results containing:
Facebook
Explore concepts:
Matthew Rowe
rowe
behaviour
the Open University
Social Web
disambiguation
Facebook
KMi
Settings
Search: the Web only in UK only in Ireland
SafeSearch - Off
3,080,000 results for
matthew rowe:
Show All
Wikipedia
BBC
Facebook
Matthew Rowe | Facebook
Facebook gives people the power to share and makes the world more open and connected. Matthew Rowe has 61 friends on Facebook and is a fan of 4 ...
en-gb.facebook.com/matt.d.rowe
Matthew Rowe | Facebook
Matthew Rowe is on Facebook. Join Facebook to connect with Matthew Rowe and others you may know. Facebook gives people the power to share and makes ...
en-gb.facebook.com/mattroweshow?_fb_noscript=1 - 54k

Yahoo!

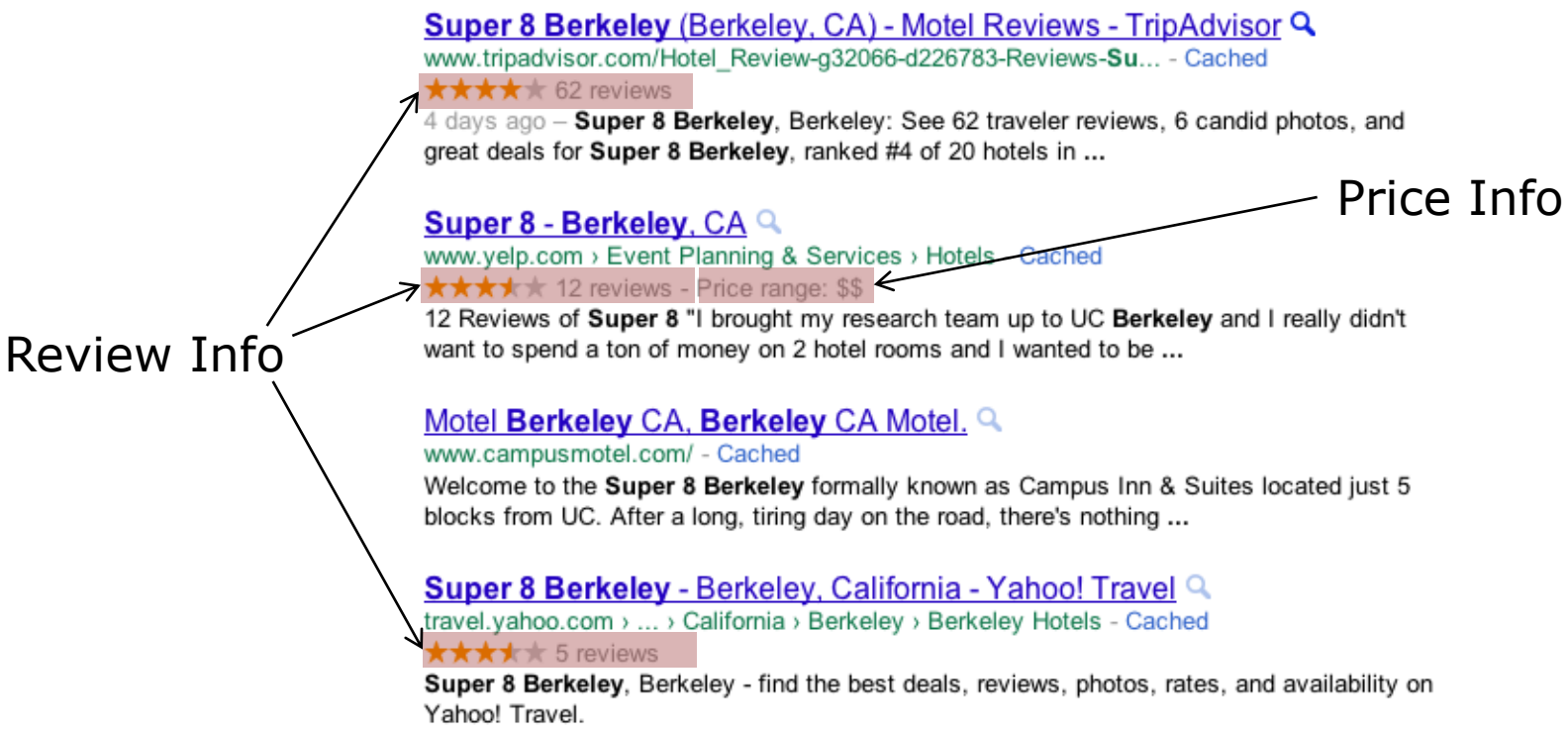


Concepts associated with the Entity

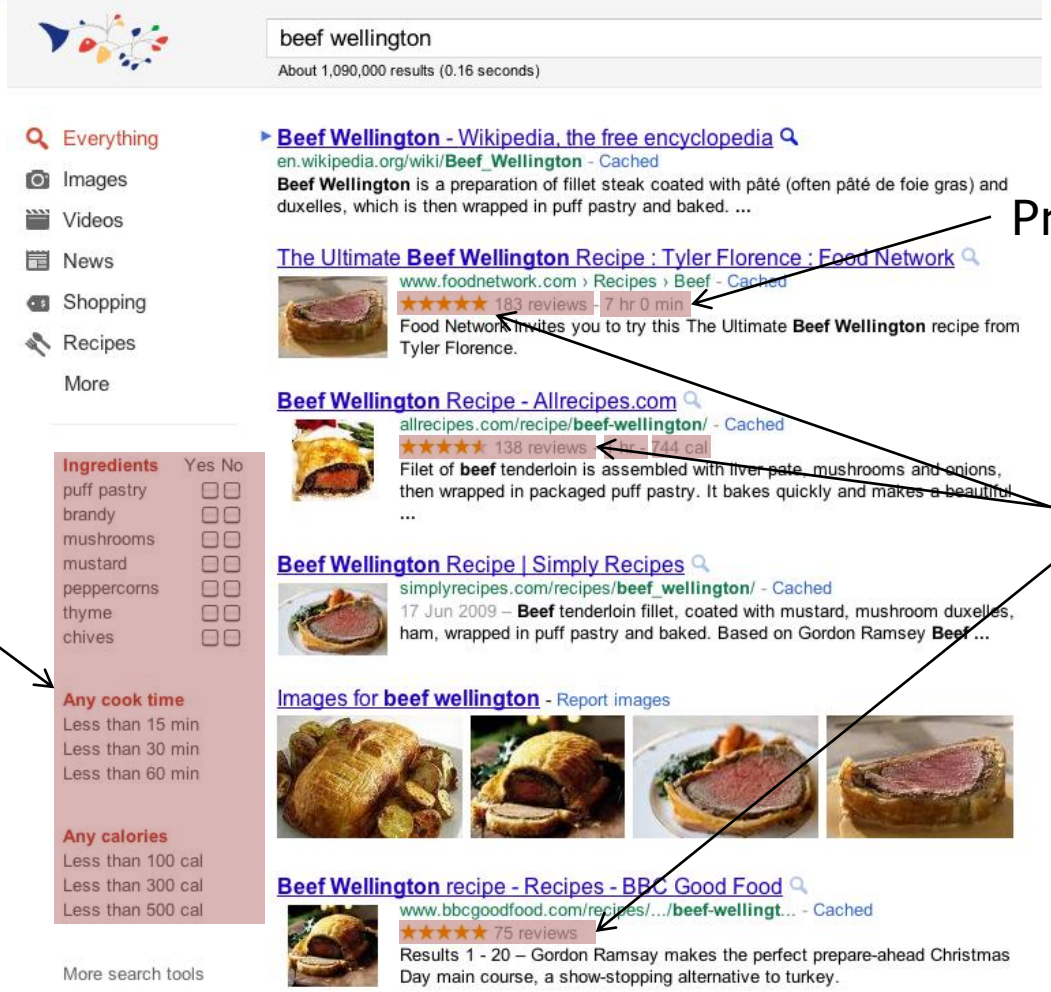


Consuming Facebook's RDFa!

Google Rich Snippets



Google Rich Snippets

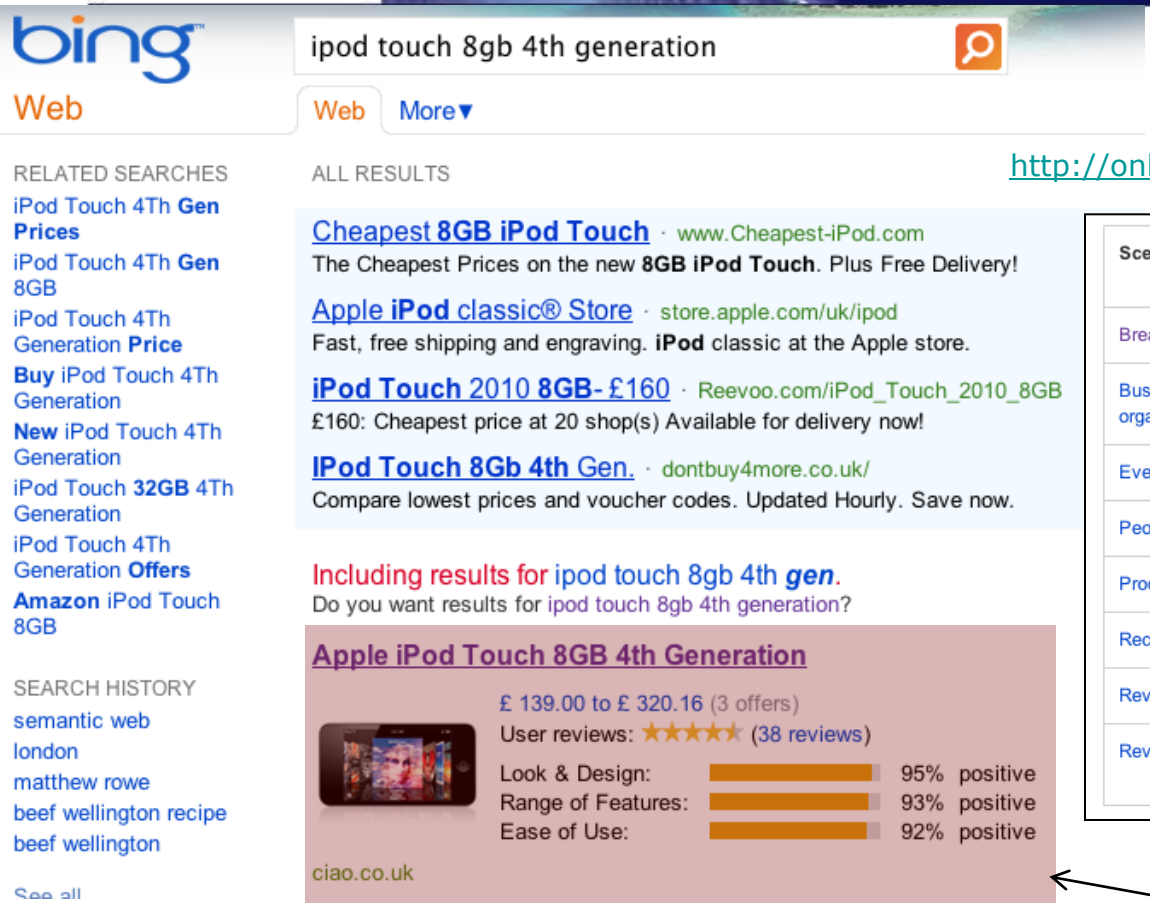


Preparation Time

Filter by attribute

Review Info

Bing



http://onlinehelp.microsoft.com/en-us/bing/hh207238.aspx

Scenario	Microdata support	Microformats support	RDFa support
Breadcrumbs	Yes	No	Yes
Businesses and organizations	Yes	Yes (hCard)	Yes
Events	Yes	Yes (hCalendar)	Yes
People	Yes	Yes (hCard)	Yes
Products and offers	Yes	Yes (hProduct)	No
Recipes	Yes	Yes (hRecipe)	Yes
Reviews	Yes	Yes (hReview)	Yes
Reviews (aggregated)	Yes	Yes (hReview-Aggregate)	Yes

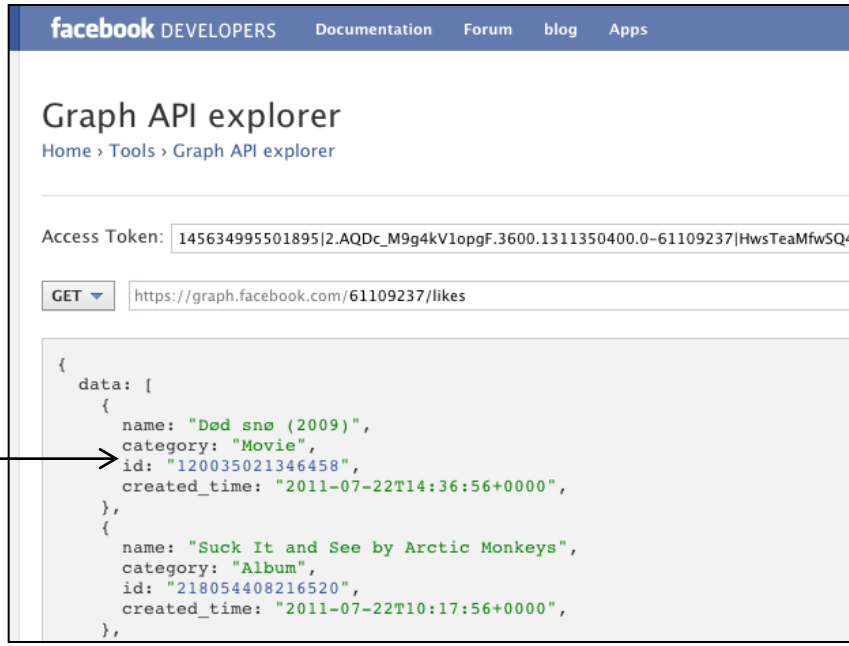
Review, price and Offer info

Enhanced User Profiles: Facebook Open Graph Protocol

- Designed to link profile pages into Facebook's social graph
  - Profile pages: movies, products, people, events, etc
  - Checkout 'Object Types': http://developers.facebook.com/docs/opengraph/
- Facebook deployed their own vocabulary of terms: Open Graph Protocol
- Web developers encode markup into page <head> element:

```
<html xmlns="http://www.w3.org/1999/xhtml"
    xmlns:og="http://ogp.me/ns#”>
 <head>
  <title>The Rock (1996)</title>
  <title>Dead Snow (2009) - IMDb</title>
  <meta property="og:url" content="http://www.imdb.com/title/tt1278340/" />
  <meta property='og:image' content='http://ia.media-imdb.com/images/M/
     MV5BMjA0NDQyMjE1MF5BMl5BanBnXkFtZTcwODE1NzY5MQ@@._V1._SX98_SY140_.jpg'>
  <meta property='og:type' content='movie' />
  <meta property='og:title' content='Dead Snow (2009)' />
  <meta property='og:site_name' content='IMDb' />
  <meta property='og:director' content='Tommy Wirkola' />
```

Enhanced User Profiles: Facebook Open Graph Protocol



Let’s revisit the scenario

Redux: A Modern Semantic Web Use Case

- Alice is a keen online shopper, and so are her friends
- She wants product suggestions based on what her friends are talking about
- Her software agent gathers her profile information from several Social Web platforms
- The agent analyses what her friends are talking about and like
- Liked products are found on several shopping sites
- Alice is shown the selection of products, ranked by price, and **why** they were selected

We can now do this! 2 years ago this was not possible

Conclusions

- Semantic technologies enable machine-readability
  - Information can now be processed at a massive-scale
  - RDF = metadata model
  - Ontologies/vocabularies = metadata descriptions
- Semantics enhance:
  - Product and service descriptions
  - Search
  - User profiling
- Semantic technology is now mainstream
  - Large growth in uptake and usage (e.g. Google, Facebook)
  - Google+ and +1's set to combine user profiling with semantics further

Slides online at:
http://www.matthew-rowe.com/?q=semanticdatalecture

Questions?

Dr Matthew Rowe
Research Associate, Knowledge Media Institute,
The Open University, UK

http://people.kmi.open.ac.uk/rowe
http://www.matthew-rowe.com
@mattroweshow