

Media Meets Semantic Web

How the BBC uses DBpedia and Linked Data to Make Connections

Andreas Müller

12th June, 2017

Technische Universität Berlin

TABLE OF CONTENTS

1. Problem
2. Objectives
3. Interlinking of concepts
4. Interlinking of documents
5. Content Link Tool
6. Conclusion

Problem

PROBLEM

Large amounts of BBC online content: **TEXT, AUDIO, VIDEO**

Domain specific microsites: **FOOD, GARDENING, SPORT**, etc. ...

PROBLEM

Large amounts of BBC online content: **TEXT, AUDIO, VIDEO**

Domain specific microsites: **FOOD, GARDENING, SPORT, etc. ...**

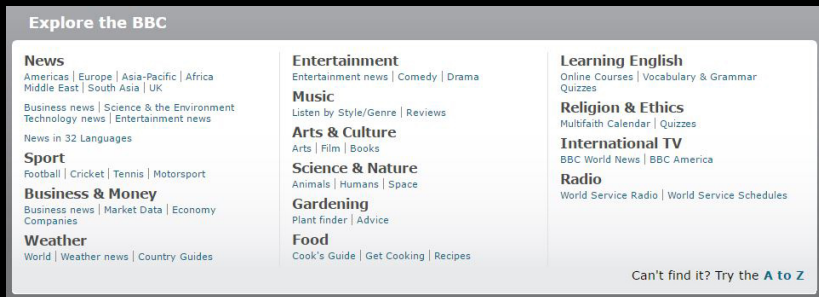


Figure 1: BBC microsites in June, 2009

PROBLEM

Large amounts of BBC online content: **TEXT, AUDIO, VIDEO**

Domain specific microsites: **FOOD, GARDENING, SPORT**, etc. ...

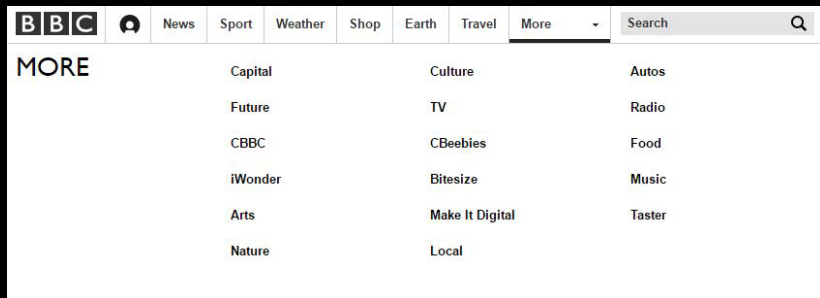


Figure 2: BBC microsites in June, 2017

PROBLEM

Not possible to...

Find **everything**, BBC has published to a given subject

Navigate between BBC domains following a **semantic thread**

No interlinking between microsites = not using full potential of available data

Objectives

OBJECTIVES

Make the BBC website **more coherent** and **more useful**

OBJECTIVES

Make the BBC website **more coherent** and **more useful**

Better connections and interlinking of existing systems

Soft transition and reducing impact on existing systems while adding new services to **maximize interlinking** of domains

OBJECTIVES

Make the BBC website **more coherent** and **more useful**

Better connections and interlinking of existing systems

Soft transition and reducing impact on existing systems while adding new services to **maximize interlinking** of domains

1. Service to link all radio and TV programmes

OBJECTIVES

Make the BBC website **more coherent** and **more useful**

Better connections and interlinking of existing systems

Soft transition and reducing impact on existing systems while adding new services to **maximize interlinking** of domains

1. Service to link all radio and TV programmes
2. Develop a new music offering

OBJECTIVES

Make the BBC website **more coherent** and **more useful**

Better connections and interlinking of existing systems

Soft transition and reducing impact on existing systems while adding new services to **maximize interlinking** of domains

1. Service to link all radio and TV programmes
2. Develop a new music offering
3. Retrofit simple navigational elements

OBJECTIVES

Make the BBC website **more coherent** and **more useful**

Better connections and interlinking of existing systems

Soft transition and reducing impact on existing systems while adding new services to **maximize interlinking** of domains

1. Service to link all radio and TV programmes
2. Develop a new music offering
3. Retrofit simple navigational elements
4. Provide a common set of web scale identifiers

OBJECTIVES

Make the BBC website **more coherent** and **more useful**

Better connections and interlinking of existing systems

Soft transition and reducing impact on existing systems while adding new services to **maximize interlinking** of domains

1. **Service to link all radio and TV programmes**
2. Develop a new music offering
3. Retrofit simple navigational elements
4. **Provide a common set of web scale identifiers**

Interlinking of concepts

Legacy auto-categorization system: CIS

Limits:

- Difficult to cover every single entity
- No relations between terms are available
- Only internal identifiers

INTERLINKING OF CONCEPTS — DBPEDIA

Common Vocabulary: DBpedia

DBpedia Label Lookup

Find most likely matches to a given term, calculate relevance with number of backlinks

Context-based Disambiguation

Disambiguate possible matches by clustering them and finding according context in DBpedia

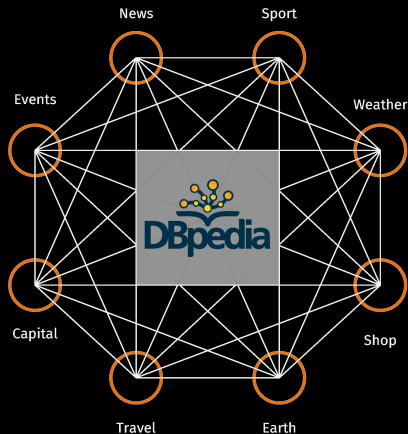


Figure 3: Linking BBC Domains

Interlinking of documents

INTERLINKING OF DOCUMENTS — MUDDY BOOTS

Identify main actors in a piece of content: **Muddy Boots**

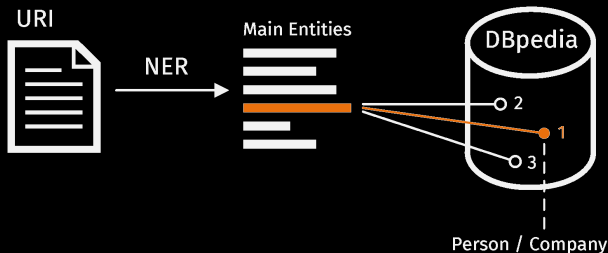


Figure 4: How Muddy Boots works

Content Link Tool

CONTENT LINK TOOL

Annotation tool to manually edit metadata

High quality automated suggestions

The screenshot displays the Content Link Tool interface with three main sections:

- Currently Added Links:** A green header with a link to remove. It contains one entry: "Arctic shrinkage".
- Machine Suggested Links:** An orange header with a link to add. It contains three entries: "Computer (magazine)", "Northern Hemisphere", and "Arctic Ocean". Below these is "Polar ice cap".
- Find another Link:** An orange header with a search prompt: "You are choosing from a controlled list of words, so here you search for the agreed term for what you mean." Below this is a search bar with the text "climate ch".

Below the search bar, a magnifying glass icon points to a search result for "Climate change". The result is highlighted in blue and includes the following text:

Climate change
Climate change is any long-term significant change in the "average weather" that a given region experiences. Average weather may include average temperature, precipitation and wind patterns. It involves changes in the variability or average state of

Below the search result, there are three more entries, each with a small icon and a title:

- Pro** (with a magnifying glass icon): Intergovernmental Panel on Climate Change
- Ed** (with a magnifying glass icon): The Intergovernmental Panel on Climate Change (IPCC) is a scientific body tasked to evaluate the risk of climate change caused by human activity. The panel was established in 1988 by the World Meteorological Organization and the United Nations Environ
- Link Te** (with a magnifying glass icon): United Nations Framework Convention on Climate Change

Below these, there is a "Blurb" section with the following text:

The United Nations Framework Convention on Climate Change (UNFCCC or FCCC) is an international environmental treaty produced at the United Nations Conference on Environment and Development, informally known as the Earth Summit, held in Rio de Janeiro

Figure 5: Suggestions of the CLT

Conclusion

CONCLUSION

User experience in the center of efforts

Smart interlinking to internal and external resources

Well integrated and hidden systems

Thanks for your attention!

Questions?



G. Kobilarov, T. Scott, Y. Raimond, S. Oliver, C. Sizemore, M. Smethurst, C. Bizer, and R. Lee.

Media meets semantic web - How the bbc uses dbpedia and linked data to make connections.

Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 5554 LNCS:723–737, 2009.