Chapter 6
Conclusions
Summary and Outlook

NIKOLAOS KONSTANTINOU
DIMITRIOS-EMMANUEL SPANOS

Outline

Introduction

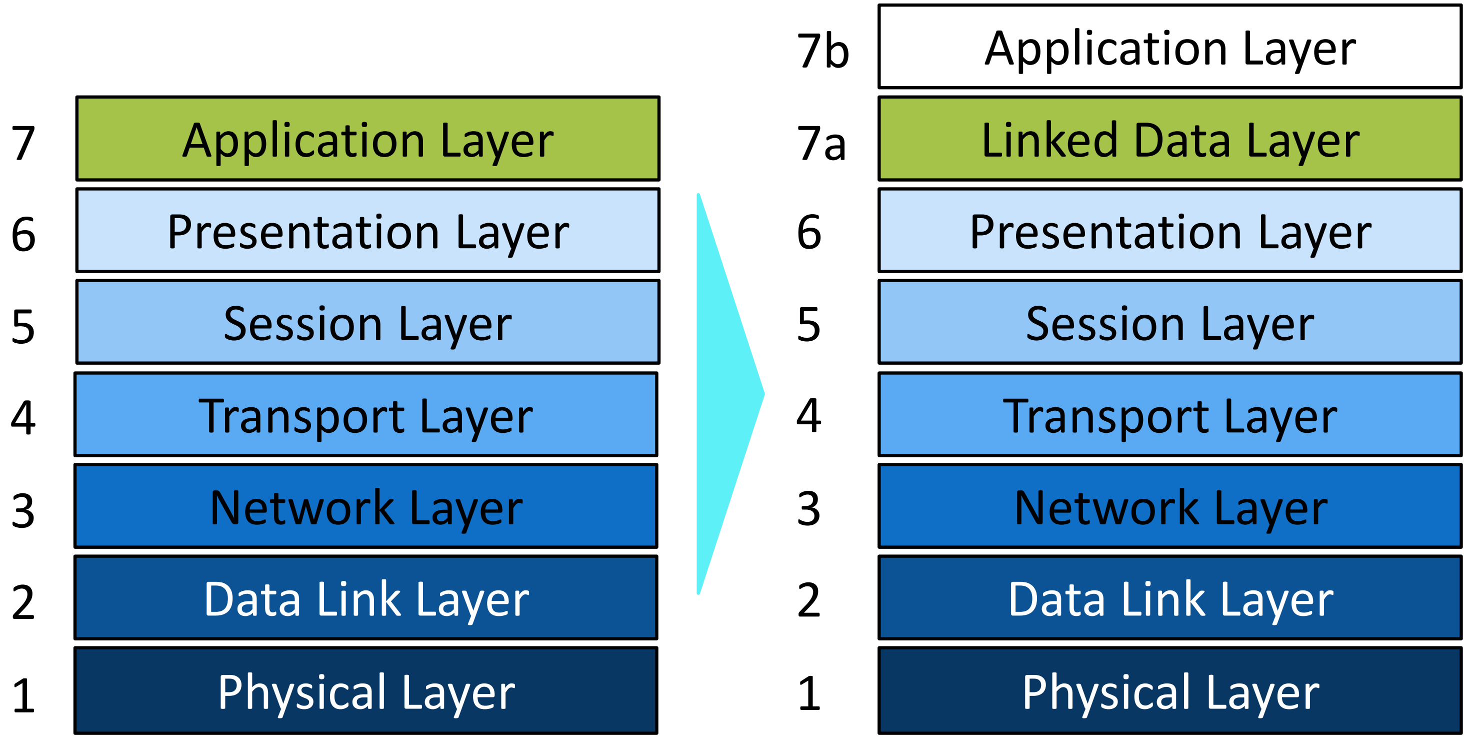
Recap

Discussion

Open Research Challenges

Introduction

An envisioned Linked Data Interoperability Layer for tomorrow's Web



Outline

Introduction

Recap

Discussion

Open Research Challenges

Chapter 1 – Introduction

Definitions – prospects – solutions

- Semantic Web
  - Main building blocks
  - Key terms
  - Issues
- Linked Data

Chapter 2 – Technical Background

From theory to practice

- Introduction of the technical background that materializes Chapter 1 concepts
- Fundamental technologies
  - From knowledge representation models to query languages and mappings
- Popular Linked Data vocabularies

Chapter 3 – Deploying Linked Data

A technical overview

Modeling data

Opening Data

Linking Data

Processing Data

- Available technical solutions and tools

Chapter 4 – Creating Linked Data from Relational Databases

RDBMS with Semantic Web applications interfaces

Motivations

Benefits

Related literature survey

- Approach categorization

Proof-of-concept use case

- Convert data from an open access repository to Linked Data

Chapter 5 – Generating Linked Data in Real-time from Sensor Data Streams

Basic concepts

- Introduction: real-time processing, context-awareness, windowing and information fusion

Related Issues

System description

- An intelligent, semantically-enabled data layer to integrate sensor information

Overall Contribution (1)

Formal and informal introduction of Data Science concepts

- Semantics
- Ontologies
- Data and Information
- Knowledge Bases
- Reasoning
- Annotation
- Metadata
- Real-time
- Context-awareness
- Integration
- Interoperability
- Fusion, etc.

Overall Contribution (2)

A detailed state-of-the-art survey

- Technologies, methodologies, tools, and approaches

Discussions on Linked Data creation

- From relational databases
- From sensor data streams

Two detailed architectural and behavioral descriptions of two domain-specific scenarios

Outline

Introduction

Recap

Discussion

Open Research Challenges

Discussion (1)

LOD ecosystem

- An open and distributed system
- Heterogeneity is inevitable
  - At a syntactic, terminological, conceptual, or semiotic/pragmatic level
- The Linked Data paradigm
  - Offers solutions to reduce heterogeneity at all four levels
    - Defines relations across the heterogeneous sources

Discussion (2)

Server-side: many steps, many components involved

- No "standard" approach
- No deterministic manner in setting behavioral priorities
- The importance of scalability
  - Volume of the produced, stored and processed information
  - The number of the data sources
  - The nature of the data
    - E.g. sensor/multimedia/social network data streams

Discussion (3)

Integration with third parties

- Semantic Web technology adoption is crucial in order to assure unambiguous definition of the information and the semantics it conveys
- E.g. as in the example in the scholarly/cultural heritage domain

Data repository turned into a Knowledge Base

- Virtually endless possibilities
  - E.g. analytical reasoning, intelligent analysis
    - Data mining, pattern extraction, etc.
  - Even in unprecedented ways

Benefits (1)

Increased discoverability

- Description of the dataset contents
- Links towards instances in other parts of the LOD cloud
- Inbound links are welcome

Reduced effort for schema modifications

- New relations allowed without modifying the database schema or contents
  - E.g. define new classes and properties

Benefits (2)

Synthesis

- Integration, fusion, mashups
- Allow searches spanning various repositories, from a single SPARQL endpoint
- Allow download of parts or the whole data

Inference

- Reasoning support
- Implicit facts can be inferred, based on the existing ones, then added to the graph

Benefits (3)

Reusability

- Third parties can reuse the data in their systems
- Including the information in their datasets, or
- By reference to the published resources

Technical Difficulties (1)

Multidisciplinarity

- Annotation task not to be underestimated
- Contributions required from several scientific domains
  - Domain-specific expertise
  - Close collaboration of the implementation team

Technology barrier

- Tools are not as mature yet as to provide guidance or warnings
- E.g. in the linking or mapping procedure
  - No design-time validity of the result

Technical Difficulties (2)

Error-prone result

- Even syntactically correct
- No automatic check of whether the concepts and properties involved are used as intended
- Errors or bad practices can go unnoticed

Concept mismatch

- Extraction of stored values into RDF not always possible
  - Identical mappings may not always be found

Technical Difficulties (3)

Exceptions to the general rule

- Automated changes will apply to the majority of the data
  - The remaining portion will require manual intervention
- Post-publishing manual interventions will be required

Open Government Data

Emerging, across governments and organizations from all over the world

- E.g. US, UK

Foster transparency, collaborative governance, innovation

Enhance citizens' quality of life through the development of novel applications

Data value decreases if not released in open formats, allowing combination and linking with other open data

Ongoing efforts to integrate open governmental data to the LOD cloud

Bibliographic Archives (1)

A huge wealth of human knowledge exists in digital libraries and open access repositories

Structured metadata

- Cataloguing, indexing, searching
- Typically trapped inside monolithic systems that support Web-unfriendly protocols for data access

Linked Data

- Allows direct reuse of the work of other librarians
  - Transforms the item-centric cataloguing to entity-based descriptions

Bibliographic Archives (2)

Several efforts at national and regional level

- E.g. Library of Congress, British National Bibliography

Linking of bibliographic data with LOD datasets from other domains

- Expected to give rise to novel applications that exploit library data in combination with other (e.g. geographical) data

Internet of Things

Billions of sensing devices currently deployed worldwide
- Already form a giant network of connected "things“
- Their number expected to continuously grow
- Uniquely identified and accessed through standard Internet protocols

Applications in diverse domains
- E.g. environmental monitoring, energy management, healthcare and home and city automation

Intelligence in IoT
- Use of ontologies and other Semantic Web technologies that support inference
- Integration of developed independently deployed IoT platforms

Outline

Introduction

Recap

Discussion

Open Research Challenges

Open Research Challenges

Data Science

> The study of the generalizable extraction of knowledge from data

LOD provision is the first step

- Not a goal in itself; a means to an end

Consuming (as opposed to producing)

- Extract intelligence, generate additional value
- E.g. visualization, analytics, text mining, named entity recognition, etc.
- Quality assessment
  - LOD quality ranges from extensively curated datasets to crowd-sourced and extracted data of relatively low quality

Big (Linked) Data (1)

Data in several media channels

- E.g. social networks, blogs, multimedia sharing services
- Generated in growing rates
- Influences professional and personal decisions and actions of individuals

Size of the dataset is a part of the problem itself

- Millions or billions of facts
- Difficult to process using conventional data processing applications
- Storage and querying problems

Big (Linked) Data (2)

No formal definition of what exactly Big Data is (and what is not)

- Commonly characterized by different properties
- All V’s for some mysterious reason
  - Volume
  - Velocity
  - Variety
  - Value
  - Veracity

Big (Linked) Data (3)

Need for

- Timely, accurate, efficient analysis of Big Data volumes
- Managing, querying and consuming
- Handling the vast amounts of data to be generated in the near future

Linked Data

- Part of the Big Data landscape
- Ideal testbed for researching key Big Data challenges

(Even) More Research Challenges

Privacy

Legal aspects

Integration and reconciliation from diverse data sources