



Dimitris Kontokostas

Address: Pierion 33A, 59100, Veria, Greece (GR)

Tel: +306945796255 (m)

Email: jimkont@gmail.com

website: <http://kontokostas.com>

Personal statement

Dimitris is a Senior Data & Knowledge Engineer at GeoPhy where he designs & implements solutions on data quality & large-scale knowledge graphs pipelines. Dimitris (almost) has a Dr.-Ing PhD on large scale knowledge extraction and quality assessment from Leipzig University (AKSW Group). He served as the head of technical developments of [DBpedia](#) for three years, is the creator of [RDFUnit](#), a unit testing framework for RDF with high industry uptake, a co-editor of the W3C specification [SHACL](#) RDF constraint language, the chair of the Shape Expressions ([ShEx](#)) W3C community group and a co-author of the “[Validating RDF data](#)” book.

Web presence on: [Twitter](#), [Github](#), [LinkedIn](#), [Google Scholar](#), [SlideShare](#) and my [homepage](#)

Key Skills

- Large-Scale Knowledge Graphs, knowledge management & knowledge representation
- Data Quality advocate & editor of SHACL
- Analytical skills, algorithmic logic, Research & Development ([1400+ citations](#))
- Excellent communication skills, both written and verbal
- Proficient in Semantic Web technologies: RDF, OWL, SPARQL, Graph DBs, Triple-stores, Linked Data
- Working with Java, Scala & bash, big data, kafka, flexible with new technologies, languages
- Well-disciplined remote worker

Employment History

Senior Data & Knowledge Engineer, GeoPhy, Netherlands (remote)

(April 2017 – present)

- Incorporate data quality techniques in the GeoPhy semantic platform
- Implement data processing agents on streaming RDF data (Apache Kafka streams)
- Co-develop data ingestion and transformation pipelines

Tech Lead, DBpedia Association, Leipzig, Germany (remote)

(March 2014 – August 2017)

- Improve the data extraction process of DBpedia
- Co-coordinate the DBpedia (developer) community
- Incorporate new sources in the DBpedia data stack (i.e. [Commons](#) and [Wikidata](#))
- Improve data precision and recall through iterative processes
- Decrease the time to release time by automating and reducing many release steps
- Lead the Google Summer of Code projects since 2014
- Leading various small & large scale projects i.e. DBpedia Live, Distributed Extraction, mapping extraction
- Always trying to achieve a lot with very few (human) resources

Research Fellow, University of Leipzig, Germany (remote)

(July 2015 – March 2017)

- Leading and coordinating the ALIGNED project on behalf of Leipzig University
- Leading the Data Quality Engineering work package for ALIGNED
- Developing tools & methods for data quality verification, deployed in production systems
- Co-designing the data quality UCR for the ALIGNED consortium
- Research on data quality, knowledge graphs, mapping alignment and data fusion
- EU funding proposal writing

RDF Data Shapes Working Group, W3C

(November 2014 – present)

- Active group member since November 2014 (joined as an invited expert)
- Involved in all discussions & teleconferences for the design of SHACL, a new data constraint language for RDF Graphs and responsible for many key design decision.
- Editor of the W3C specification for [SHACL](#)

Researcher / Knowledge Engineer, University of Leipzig, Germany (remote)

(Sep 2012 – July 2015)

- PhD in large-scale knowledge extraction and quality assessment
- Bootstrapped the Dutch DBpedia, funded by the National Libraries of Netherlands
- Implementation of RDFUnit, a popular Java RDF unit testing framework with high industry uptake
- Worked on improving DBpedia wrt increasing data precision & recall
- Authored many research paper (800+ citations, h-index: 10, i10-index: 10 from [Google Scholar](#))

R&D Software Engineer (part-time), [Semantic Web Unit](#), Aristotle Un.(Remote)

April 2012 - August 2012

- UCR for utilizing the wealth of electronic data of the Aristotle University of Thessaloniki (A.U.Th.) within the global Semantic Web (Web3.0) and Linked Data..
- Drupal Developer

R&D Software Engineer VC++ (part-time), [AIIA Lab](#), Aristotle Univ. (Remote)

(February 2005 - June 2006)

- Graphical application for visualizing and editing MPEG-7 XML files with VC++ 6.0/MFC
- Video Player Implemented with DirectX featuring video annotation overlays, other Visual Components such as timelines, property trees.

Software Engineer C++, [TESAE](#), Athens, Greece (Remote)

(September 2014 - January 2005)

- Implementation of the pegasus language for UNIX using FLEX / BISON and dBase open source software (XBase).

ICT Teacher, Ministry of Education, Greece

(September 2006 - September 2011)

ICT Teacher @ Greek Secondary Education,.

R&D Software Engineer (part-time), [MUSIC Lab](#), Technical Uni. of Crete

(June 2001 - July 2003)

- Implementation of XML Parsers and XML Databases (storing and retrieving XML Documents using relational databases) in DOM/SAX using both Java and C++.

Education

PhD in Semantic Web, University of Leipzig, Leipzig, Germany

(September 2012 – present)

PhD on *large-scale knowledge extraction and quality assessment* (in progress)

MSc in Web Science, Aristotle University of Thessaloniki, Greece

(September 2009 – May 2012)

Master's degree in Web Science. A 3-semester curriculum with Thesis. Thesis entitled: “Towards Linked Data Internationalization: The successful case of Greek DBpedia”.

Grade: 9.2 (Scale 0 - 10)

Diploma in Electronic & Computer Engineering, Technical University of Crete, Greece

(September 1999 – July 2004)

Bachelor/Diploma degree in Electronic & Computer Engineering. Technical University of Crete. A 5-year curriculum with thesis. Thesis entitled “Grid resource Brokering”

Grade: 8.26 (Scale 0 – 10)

Projects

RDFUnit

<https://github.com/AKSW/RDFUnit>

Popular open source Java RDF Unit Testing suite with an industry uptake. Main developer

DBpedia extraction framework

<https://github.com/DBpedia/extraction-framework> , <https://github.com/DBpedia>

Structured data (RDF) extraction of Wikipedia. Technical & technology Lead

ALIGNED Software & Data Engineering

<http://aligned-project.eu/> (02.2015 - 01.2018)

ALIGNED is an EU2020 funded project that tries to align software & data engineering lifecycles and provide tools and architectures for greater agility in data intensive systems. I am leading work package 4, regarding data quality engineering and coordinate the project on behalf of Leipzig University

DBpedia @ GSoC 2013, 2014, 2015, 2016 (& 2017 in progress)

Project administrator and mentor of popular DBpedia projects like commons.dbpedia.org, wikidata.dbpedia.org, the new Linked Data interface and the Spark-based distributed extraction.

[GSoC2013](#), [GSoC2014](#), [GSoC2015](#) , [GSoC2016](#)

Skills

- **Languages:** Java / Scala (Very good), C / C++ (rusty) , shell script, PHP (good), C#, Delphi, JSP, ASP (familiar), SQL (Excellent)
- **Web Standards:** Linked Data, RDF, RDFS, OWL, SPARQL, XML, XML Schema, HTML, CSS
- **Databases:** Virtuoso (RDF), MySQL, Postgres, SQL Server
- **Server:** Apache, Tomcat, NGinx
- **OS:** Linux (Ubuntu, Centos)
- **Streaming / stream processing:** Kafka, Spark

Languages

- **English:** Excellent
- **Greek:** Excellent (native)