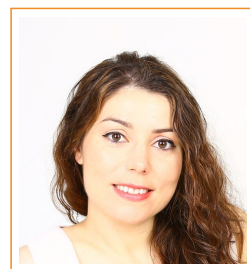


# Marilena Oita

Via Tesserete 7  
6900 Massagno, Swizerland  
☎ +41 (0)76 432 987 4  
✉ marilena@idsia.ch  
Date of birth: 24.09.1985  
[Linkedin profile](#)



## Professional experience

- 12/2016–current **Researcher, IDSIA.**
- research focus: text mining and information extraction
  - projects: Clabs, DufEnergy
  - participate to teaching activities (Big Data with Spark)
- 05/2016–12/2016 **Lead Data Engineer, ManoMano.**
- structure data warehouse in order to make data actionable
  - produce and analyze customer/provider behavior KPIs
  - drive growth through analytics
  - build internal models using machine learning algorithms
- 06/2014–05/2016 **Machine Learning Solution Architect, CustomerMatrix.**
- data pre-processing, features engineering, semantic data analysis and enrichment
  - build predictive models for customer profiling, churn prevention, product & action recommendation
  - machine learning algorithms for clustering, classification, and recommendation
  - tools: Weka, Spark (core, MLLib & GraphX);
  - Solr-based platform for data indexing;
- 12/2012–04/2014 **R&D Engineer, Internet Memory Research, (IMR), France.**
- implement the text mining application level of the platform
  - semantic features and KPIs extraction
  - topic matching and categorisation of Web pages
  - tools: Apache Lucene, Hadoop and Hbase(cloudera dist), ElasticSearch;
- 10/2009–11/2012 **PhD, Télécom ParisTech, Paris, France.**
- topic: *Deriving Semantic Objects from the Structured Web*
  - part of the Webdam Project
  - implementation of systems and data analysis for: main content extraction from Web pages [4, 2]; Deep Web exploration using Information Extraction and Ontology Matching [1]; Web pages dynamics [3]
- 2010–2012 **Teaching Assistant.**
- weekly laboratory classes*
- *Web Search technologies 2012, @Telecom ParisTech*
  - *Databases 2012, @ESIEA*
  - *C Programming Language 2010, @UVSQ*
- 02/2008–08/2008 **Intern, INRIA, Saclay, Ile-de-France, France.**
- 03/2009–07/2009 **Topic: Community systems in Peer-to-Peer**
- elaborate the model of a secure distributed social network based on ActiveXML
  - implement the interface of social network using J2EE

---

## Education

- 10/2004–07/2009 **Bachelor in Computer Science and Engineering**, *University 'Politehnica'*, of Bucharest, Romania.  
top 5%
- 09/2007–08/2008 **Master1**, *Informatique École Polytechnique(X)*, Palaiseau, France.
- 09/2003–07/2007 **Baccalaureate degree**, '*Callatis*', Mangalia, Romania.  
math-info

---

## Skills

- Algorithms advanced
- Programming *Java, Scala, R*
- Machine Learning Spark (ML and GraphX), Weka
- Development *Eclipse/Netbeans, Maven, Mercurial/SVN*
- Text Mining *Natural Language Processing Information Extraction Information Retrieval*
- Web J2EE, JAXB, JAX-RS based REST API implementation
- Semantic Web ontology exploitation(*YAGO*), entity disambiguation(*AIDA, WordNet*), concepts extraction
- Databases SQL *Postgres*, NoSQL *Hbase*
- Linux Shell scripting
- Other  $\LaTeX$

---

## Languages

- English advanced *fluent*
- French advanced *fluent*
- Italian beginner *good understanding*
- Romanian native *mother tongue*

---

## Publications

- [1] Marilena Oita, Antoine Amarilli, and Pierre Senellart. Cross-fertilizing deep Web analysis and ontology enrichment. In *Proc. Very Large Data Search VLDS*, 2012.
- [2] Marilena Oita and Pierre Senellart. Archiving data objects using Web feeds. In *Proc. International Workshop on Web archiving IAWW*, 2010.
- [3] Marilena Oita and Pierre Senellart. Deriving dynamics of Web pages: A survey. In *Proc. Temporal Workshop on Web archiving TWAW*, 2011.
- [4] Marilena Oita and Pierre Senellart. FOREST: Focused object retrieval by exploiting significant tag paths. In *preparation submission to IEEE Transactions on Knowledge and Data Engineering TKDE Journal*, 2013.