florian **leitner**

data scientist

contact

Address Line
E-280XX Madrid

+34 6XX XXX XXX
florian.leitner@gmail.com

home://fnl.es
work://github.com/fnl

languages

German: native speaker

English: mother tongue

Spanish: fluent (s/r/w)

computing

probabilistic programming

natural language processing

database administration

web (service) development

software engineering

Python, Scala, R, SQL, C, Perl, Java, …

profile

Bioinformatics and *molecular biology*.
Machine learning for *data and text mining*.
Query engine and *information extraction* development.
Distributed *services* and parallel *high-performance computing*.
Excellent *communication, planning, and organization* skills.

experience

2014–now   **Technical University of Madrid (UPM)**   Madrid, Spain
Juan de la Cierva research fellow (Data mining software development)
Development of an integrated text mining/information extraction system for neuro-biology with Scala (Factorie, Hadoop/Spark), Python (brat, SciKit-Learn) and R for data visualization. Teaching machine learning and text mining to students.

2013–2014   **National Cancer Research Centre (CNIO)**   Madrid, Spain
Post-doc (Data mining & NLP software development)
Development of a large-scale text mining/information extraction pipeline in Java (UIMA), Scala (Hadoop/Spark), C (var. machine learning libraries) and Python (SciKit-Learn) for genome-wide transcription regulation interactions.

2006–2013   **National Cancer Research Centre (CNIO)**   Madrid, Spain
PhD student (Bioinformatics software development)
Designer and lead of a dozen international research teams that implemented the BioCreative Meta-Server (Python/Django, Postgres, Java). Co- and main organizer of large, international scientific community challenges: BioCreative II (co), II.5 (main), III (co), and IV (co). Development of the BioCreative web site (www.biocreative.org) and team management interface (Python/Django, Postgres). Exploratory data analyses and evaluations in R.

2003–2006   **Institute for Molecular Pathology (IMP)**   Vienna, Austria
MSc student (Database administration and software/web development)
Data acquisition and import from biological resources (e.g., parser development) in Perl and LISP; Web-based data visualization (Apache Struts) for an in-hose bioin-formatics framework ("Annotator"). Design and implementation of an indexing engine and the query interface with Apache Lucene.

2003–2003   **CSO Lanifex GmbH**   Vienna, Austria
Consultant (Linux kernel and Firefox browser customization)
Development of a specialized VPN network for the Austrian National Bank (Linux kernel modifications and adapting the Mozilla framework as a GUI for the thin clients). Contributions to the Vienna United Nations intranet web-site (MySQL).

2002–2003   **European Media Labs (EML)**   Heidelberg, Germany
Python/Java application developer
Development of a 3D visualization plug-in for annotations of protein structures (Java). Installation and administration of Python libraries for Structural Biology.

2001–2002   **Novartis Research Institute**   Vienna, Austria
Web developer and database administration
Development of a Perl-based bioinformatics framework for a large Alpha cluster.

education

| 2012 | **PhD Bioinformatics, with distinction** | Autonomous University of Madrid (UAM), Spain |

Online assessment of protein interaction information extraction systems.

| 2006 | **MSc Molecular Biology, with distinction** | Vienna University, Austria |

Integration of biological databases and mining for post-translational modifications.

| 1995 | **Advertisement & Design School** | Innsbruck, Austria |

Werbe- und Design Akademie

| 1994 | **Business High-school** | Innsbruck, Austria |

Handelsschule Innsbruck

accomplishments

| 2014 | **Machine Learning** | Stanford/Coursera MOOC |

completed with 100+/100 total score

| 2013 | **Special Award - Premio Extraordinario** | Autonomous University of Madrid (UAM) |

for my PhD thesis work

| 2012 | **Speed Reading** | Programa Integral de Lectura |

accredited final reading speed: 2000 words/minute

| 2012 | **Natural Language Processing** | Stanford/Coursera MOOC |

completed with 88/100 total score

| 2009 | **Main Conference Organizer** | BioCreative, CNIO, Madrid |

three-day conference with nearly 100 participants

| 2009 | **Services Cup** | IEEE Services-I World Conference |

3rd place behind competitors from SAP and IBM

communication skills

| 2014 | **Invited Talk** | Institute for Computational Linguistics (UZH), Zurich, Switzerland |

From the BioCreative challenges to transcription regulation network extraction.

| 2014 | **Progress Report** | National Cancer Research Centre (CNIO), Madrid, Spain |

Retracing transcription regulation networks.

| 2013 | **Presentation** | BioLINK @ Intelligent Systems for Molecular Biology (ISMB), Berlin, Germany |

Mining cis-regulatory transcription networks from literature.

| 2013 | **Panel Host** | BioLINK @ Intelligent Systems for Molecular Biology (ISMB), Vienna, Austria |

Integration across genomics and medicine.

| 2011 | **Invited Talk** | FEBS Journal Board Meeting, Cambridge, UK |

Annotating scientific manuscripts and the role of text-mining.

| 2010 | **Progress Report** | National Cancer Research Centre (CNIO), Madrid, Spain |

Making biological information accessible.

| 2010 | **Presentation** | Intelligent Systems for Molecular Biology (ISMB), Stockholm, Sweden |

Comparative community assessments for applied biomedical text mining.

| 2009 | **Presentation** | Semantic Mining in Biomedicine (SMBM), Turku, Finland |

The BioCreative Meta-Server platform.

| 2008–2013 | **Presentations** | BioCreative, Madrid, Spain and Bethesda, MD, USA |

Various presentations throughout the BC II-IV workshops (see conference proceedings).

publications

journal articles

BioC: a minimalist approach to interoperability for biomedical text processing
DC Comeau, R Islamaj Dogan, P Ciccarese, KB Cohen, M Krallinger, F Leitner, ... WJ Wilbur
Database *2013 (Sept. 2013) bat064–bat064. 2013*

MyMiner: a web application for computer-assisted biocuration and text annotation
D Salgado, M Krallinger, M Depaule, E Drula, A Tendulkar, F Leitner, A Valencia, C Marcelle
Bioinformatics *(2012) bts435. 2012*

How to link ontologies and protein-protein interactions to literature: text-mining approaches and the BioCreative experience.
M Krallinger, F Leitner, M Vazquez, D Salgado, C Marcelle, M Tyers, A Valencia, A Chatr-Aryamontri
Database *2012 (2012) bas017. 2012*

Text Mining for Drugs and Chemical Compounds: Methods, Tools and Applications
M. Vazquez, M Krallinger, F Leitner, A Valencia
Molecular Informatics *30.6-7 (2011) pp. 506–519. 2011*

The Protein-Protein Interaction tasks of BioCreative III: classification/ranking of articles and linking bio-ontology concepts to full text
Martin Krallinger, Miguel Vazquez, Florian Leitner, ... Alfonso Valencia
BMC Bioinformatics *12.Suppl 8 (2011) S3. 2011*

The FEBS Letters SDA corpus.
Florian Leitner, Martin Krallinger, Gianni Cesareni, Alfonso Valencia
FEBS Letters *584.19 (Oct. 2010) pp. 4129–4130. 2010*

An Overview of BioCreative II.5.
Florian Leitner, Scott A Mardis, Martin Krallinger, Gianni Cesareni, Lynette A Hirschman, Alfonso Valencia
IEEE/ACM Transactions on Computational Biology and Bioinformatics *7.3 (June 2010) pp. 385–399. 2010*

The PPI affix dictionary (PPIAD) and BioMethod lexicon
M Krallinger, AV Tendulkar, F Leitner, A Chatr-aryamontri, A Valencia
BMC Bioinformatics *11 (2010) pp. 1–3. 2010*

The FEBS Letters/BioCreative II.5 experiment: making biological information accessible
F Leitner, A Chatr-Aryamontri, SA Mardis, A Ceol, M Krallinger, L Licata, L Hirschman, G Cesareni, A Valencia
Nature Biotechnology *28.9 (2010) pp. 897–899. 2010*

A text-mining perspective on the requirements for electronically annotated abstracts.
Florian Leitner, Alfonso Valencia
FEBS Letters *582.8 (Apr. 2008) pp. 1178–1181. 2008*

Overview of the protein-protein interaction annotation extraction task of BioCreative II.
Martin Krallinger, Florian Leitner, Carlos Rodriguez-Penagos, Alfonso Valencia
Genome Biology *9 Suppl 2 (2008) S4. 2008*

Evaluation of text-mining systems for biology: overview of the Second BioCreative community challenge.
M Krallinger, A Morgan, L Smith, F Leitner, L Tanabe, J Wilbur, L Hirschman, A Valencia
Genome Biology *9 Suppl 2 (2008) S1. 2008*

Introducing meta-services for biomedical information extraction.
F Leitner, M Krallinger, C Rodriguez-Penagos, J Hakenberg, ... C Sander, A Valencia
Genome Biology *9 Suppl 2 (2008) S6. 2008*

ProSAT: functional annotation of protein 3D structures.
Razif R Gabdoulline, René Hoffmann, Florian Leitner, Rebecca C Wade
Bioinformatics *19.13 (Sept. 2003) pp. 1723–1725. 2003*

book chapters

BioCreative II.5 and the FEBS Letters Experiment on Structured Digital Abstracts
Florian Leitner, Martin Krallinger, Valencia Alfonso
Encyclopedia of Systems Biology pp. 101–106, 2013, Springer New York

BioCreative Meta-Server and Text-Mining Interoperability Standard
Florian Leitner, Martin Krallinger, Valencia Alfonso
Encyclopedia of Systems Biology pp. 106–110, 2013, Springer New York

Analysis of Biological Processes and Diseases Using Text Mining Approaches
Martin Krallinger, Florian Leitner, Alfonso Valencia
Methods in Molecular Biology pp. 341–382, 2010, Humana Press

Biological Knowledge Extraction
Florian Leitner, Robert Hoffmann, Alfonso Valencia
Bioinformatics for Systems Biology pp. 413–433, 2009, Humana Press

conference proceedings

Retrieval and Discovery of Cell Cycle Literature and Proteins by Means of Machine Learning, Text Mining and Network Analysis
Martin Krallinger, Florian Leitner, Alfonso Valencia
8th International Conference on Practical Applications of Computational Biology & Bioinformatics, 2014

Overview of the chemical compound and drug name recognition (CHEMDNER) task
Martin Krallinger, Florian Leitner, Obdulia Rabal, Miguel Vazquez, Julen Oyarzabal, Alfonso Valencia
BioCreative IV Challenge Evaluation Workshop, 2013

Mining cis-Regulatory Transcription Networks from Literature
Florian Leitner, Martin Krallinger, Sushil Tripathi, Martin Kuiper, Astrid Lćgreid, Alfonso Valencia
Proceedings of the ISMB BioLINK SIG, 2013

Results of the BioCreative III Interaction Method Task
Martin Krallinger, Miguel Vazquez, Florian Leitner, Alfonso Valencia
Proceedings of the BioCreative III Challenge Evaluation Workshop, 2010

The BioCreative II.5 challenge overview
Martin Krallinger, Florian Leitner, Alfonso Valencia
Proceedings of the BioCreative II.5 Challenge Evaluation Workshop, 2009

SOA-Based Integration of Text Mining Services
Johann Starlinger, Florian Leitner, Alfonso Valencia, Ulf Leser
Proceedings of the 2009 World Conference on Services, 2009

Exploiting text mining of biological annotations to extract protein interactions for the Semantic web
Martin Krallinger, Carlos Rodriguez-Penagos, Florian Leitner, Alfonso Valencia
Seventh International Workshop on Network Tools and Applications in Biology, 2007

Assessment of the second BioCreative PPI task: automatic extraction of protein-protein interactions
Martin Krallinger, Florian Leitner, Alfonso Valencia
Proceedings of the second BioCreative Challenge Evaluation Workshop, 2007