

florian leitner

data scientist

contact

Address Line
E-280XX Madrid

+34 6XX XXX XXX
leitner@
datacatalytics.com

[http://
datacatalytics.com](http://datacatalytics.com)

languages

German: native

English: master

Spanish: proficient

computing

scientific computing: R,
SciPy, GNU/POSIX/C,
Scala

data engineering:
PostgreSQL, HDFS,
Cassandra

language processing:
StanfordNLP, Factorie,
LingPipe, etc.

distributed computing:
UIMA, Spark,
ElasticSearch

software development:
JetBrains IDEs,
Jenkins/Travis,
Gradle/Maven, Ansible,
etc.

profile

Bioinformatics and molecular biology expert.
Statistical learning for data and text mining.
Data engineering and information extraction.
Distributed and parallel high-performance computing.
Excellent communication, planning, and organization skills.

experience

- | | | |
|-----------|--|---------------------|
| 2015–now | Leitner Catalytics, S.L.
International consultant (Data mining & NLP)
Delivering tailored, cloud-based machine learning and text mining solutions to clients world-wide (in Java/Scala, Python, and R). | Madrid, Spain |
| 2014–2015 | Technical University of Madrid (UPM)
Juan de la Cierva research fellow (Inference engine & NLP development)
Information extraction for neurobiology with Factorie, Spark, SciKit-Learn, OrientDB, and R. Statistical analyses (R). Teaching (R, Python, GATE). | Madrid, Spain |
| 2013–2014 | National Cancer Research Centre (CNIO)
Post-doc (Data mining & NLP software development)
Extraction of the mammalian transcriptome with UIMA, Lucene, C/C++, SciKit-Learn, and D3.js. | Madrid, Spain |
| 2006–2013 | National Cancer Research Centre (CNIO)
PhD student (Bioinformatics software development)
Co- and main organizer of BioCreative (BC). Lead devel. of the BC meta-server with a dozen intl. teams (Python, Postgres, Java). BioCreative web site and team management interface (Django, Postgres). Analyses and evaluations in R. | Madrid, Spain |
| 2003–2006 | Institute for Molecular Pathology (IMP)
MSc student (Database administration and software/web development)
Data acquisition from biological resources (Perl, LISP); Web visualization (Struts) of those resources. Indexing system and query interface (Lucene). | Vienna, Austria |
| 2003–2003 | CSO Lanifex GmbH
Consultant (Linux kernel and Firefox browser customization, web dev.)
Contributions to a VPN network for the Austrian National Bank (Linux kernel and Firefox mods) and the United Nations intranet (PHP, MySQL). | Vienna, Austria |
| 2002–2003 | European Media Labs (EML)
Python/Java application developer
3D viz. of protein structure annotations (Java). Structural biology (Python). | Heidelberg, Germany |
| 2001–2002 | Novartis Research Institute
Web developer and database administration
Perl-based bioinformatics framework development for a large Alpha cluster. | Vienna, Austria |
| 1997–2000 | Pro-Design Paragliders
Global sales manager
Overseas production and intl. sales. Stock management system (VBasic). | Innsbruck, Austria |

education

2012	PhD Bioinformatics, with distinction Online assessment of protein interaction information extraction systems.	Autonomous University of Madrid (UAM), Spain
2006	MSc Molecular Biology, with distinction Integration of biological databases and mining for post-translational modifications.	Vienna University, Austria

accomplishments

2015	Apache Spark Big Data and Scalable Machine Learning both completed with 100/100 total scores	edX Berkeley MOOC
2014	Machine Learning completed with 100+/100 total score	Stanford/Coursera MOOC
2013	Special Award - Premio Extraordinario for my PhD thesis work	Autonomous University of Madrid (UAM)
2012	Speed Reading accredited final reading speed: 2000 words/minute	Programa Integral de Lectura
2012	Natural Language Processing completed with 88/100 total score	Stanford/Coursera MOOC
2009	Main Conference Organizer three-day conference with nearly 100 participants	BioCreative, CNIO, Madrid
2009	Services Cup 3 rd place behind competitors from SAP and IBM	IEEE Services-I World Conference

communication skills

2015	Invited Talk The BioCreative challenges: Critical assessments of biomedical information extraction.	Jornada de Tecnología y Salud, Real Academia de Ingeniería, Madrid, Spain
2014	Invited Talk From the BioCreative challenges to transcription regulation network extraction.	Institute for Computational Linguistics (UZH), Zurich, Switzerland
2014	Progress Report Retracing transcription regulation networks.	National Cancer Research Centre (CNIO), Madrid, Spain
2013	Presentation Mining cis-regulatory transcription networks from literature.	BioLINK @ Intelligent Systems for Molecular Biology (ISMB), Berlin, Germany
2013	Panel Host Integration across genomics and medicine.	BioLINK @ Intelligent Systems for Molecular Biology (ISMB), Vienna, Austria
2011	Invited Talk Annotating scientific manuscripts and the role of text-mining.	FEBS Journal Board Meeting, Cambridge, UK
2010	Progress Report Making biological information accessible.	National Cancer Research Centre (CNIO), Madrid, Spain
2010	Presentation Comparative community assessments for applied biomedical text mining.	Intelligent Systems for Molecular Biology (ISMB), Stockholm, Sweden
2009	Presentation The BioCreative Meta-Server platform.	Semantic Mining in Biomedicine (SMBM), Turku, Finland
2008–2013	Presentations Various presentations throughout the BC II-IV workshops (see conference proceedings).	BioCreative, Madrid, Spain and Bethesda, MD, USA

publications

journal articles

CHEMDNER: The drugs and chemical names extraction challenge

Krallinger, M, F Leitner, O Rabal, M Vazquez, J Oyarzabal, and A Valencia

Journal of Cheminformatics 7.Suppl 1 (2015) S1. 2015

The CHEMDNER corpus of chemicals and drugs and its annotation principles

Krallinger, M, O Rabal, F Leitner, M Vazquez, D Salgado, ..., J Oyarzabal, and A Valencia

Journal of Cheminformatics 7.Suppl 1 (2015) S2. 2015

BioC: a minimalist approach to interoperability for biomedical text processing

Comeau, DC, R Islamaj Dogan, P Ciccarese, KB Cohen, M Krallinger, F Leitner, ..., and WJ Wilbur

Database 2013 (Sept. 2013) bat064–bat064. 2013

How to link ontologies and protein-protein interactions to literature: text-mining approaches and the BioCreative experience.

Krallinger, M, F Leitner, M Vazquez, D Salgado, C Marcelle, M Tyers, A Valencia, and A Chatr-Aryamontri

Database 2012 (2012) bas017. 2012

MyMiner: a web application for computer-assisted biocuration and text annotation

Salgado, D, M Krallinger, M Depaule, E Drula, A Tendulkar, F Leitner, A Valencia, and C Marcelle

Bioinformatics (2012) bts435. 2012

The Protein-Protein Interaction tasks of BioCreative III: classification/ranking of articles and linking bio-ontology concepts to full text

Krallinger, Martin, Miguel Vazquez, Florian Leitner, ..., and Alfonso Valencia

BMC Bioinformatics 12.Suppl 8 (2011) S3. 2011

Text Mining for Drugs and Chemical Compounds: Methods, Tools and Applications

Vazquez, M., M Krallinger, F Leitner, and A Valencia

Molecular Informatics 30.6-7 (2011) pp. 506–519. 2011

The PPI affix dictionary (PPIAD) and BioMethod lexicon

Krallinger, M, AV Tendulkar, F Leitner, A Chatr-aryamontri, and A Valencia

BMC Bioinformatics 11 (2010) pp. 1–3. 2010

The FEBS Letters/BioCreative II.5 experiment: making biological information accessible

Leitner, F, A Chatr-Aryamontri, SA Mardis, A Ceol, M Krallinger, L Licata, L Hirschman, G Cesareni, and A Valencia

Nature Biotechnology 28.9 (2010) pp. 897–899. 2010

The FEBS Letters SDA corpus.

Leitner, Florian, Martin Krallinger, Gianni Cesareni, and Alfonso Valencia

FEBS Letters 584.19 (Oct. 2010) pp. 4129–4130. 2010

An Overview of BioCreative II.5.

Leitner, Florian, Scott A Mardis, Martin Krallinger, Gianni Cesareni, Lynette A Hirschman, and Alfonso Valencia

IEEE/ACM Transactions on Computational Biology and Bioinformatics 7.3 (June 2010) pp. 385–399. 2010

Evaluation of text-mining systems for biology: overview of the Second BioCreative community challenge.

Krallinger, M, A Morgan, L Smith, F Leitner, L Tanabe, J Wilbur, L Hirschman, and A Valencia

Genome Biology 9 Suppl 2 (2008) S1. 2008

Overview of the protein-protein interaction annotation extraction task of BioCreative II.

Krallinger, Martin, Florian Leitner, Carlos Rodriguez-Penagos, and Alfonso Valencia

Genome Biology 9 Suppl 2 (2008) S4. 2008

Introducing meta-services for biomedical information extraction.

Leitner, F, M Krallinger, C Rodriguez-Penagos, J Hakenberg, ..., C Sander, and A Valencia

Genome Biology 9 Suppl 2 (2008) S6. 2008

A text-mining perspective on the requirements for electronically annotated abstracts.

Leitner, Florian and Alfonso Valencia

FEBS Letters 582.8 (Apr. 2008) pp. 1178–1181. 2008

ProSAT: functional annotation of protein 3D structures.

Gabdoulline, Razif R, René Hoffmann, Florian Leitner, and 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

Leitner, Florian, Martin Krallinger, and Valencia Alfonso

Encyclopedia of Systems Biology pp. 101–106, 2013, Springer New York

BioCreative Meta-Server and Text-Mining Interoperability Standard

Leitner, Florian, Martin Krallinger, and Valencia Alfonso

Encyclopedia of Systems Biology pp. 106–110, 2013, Springer New York

Analysis of Biological Processes and Diseases Using Text Mining Approaches

Krallinger, Martin, Florian Leitner, and Alfonso Valencia

Methods in Molecular Biology pp. 341–382, 2010, Humana Press

Biological Knowledge Extraction

Leitner, Florian, Robert Hoffmann, and 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

Krallinger, Martin, Florian Leitner, and Alfonso Valencia

8th International Conference on Practical Applications of Computational Biology & Bioinformatics, 2014

Overview of the chemical compound and drug name recognition (CHEMDNER) task

Krallinger, Martin, Florian Leitner, Obdulia Rabal, Miguel Vazquez, Julen Oyarzabal, and Alfonso Valencia

BioCreative IV Challenge Evaluation Workshop, 2013

Mining cis-Regulatory Transcription Networks from Literature

Leitner, Florian, Martin Krallinger, Sushil Tripathi, Martin Kuiper, Astrid Lægreid, and Alfonso Valencia

Proceedings of the ISMB BioLINK SIG, 2013

Results of the BioCreative III Interaction Method Task

Krallinger, Martin, Miguel Vazquez, Florian Leitner, and Alfonso Valencia

Proceedings of the BioCreative III Challenge Evaluation Workshop, 2010

The BioCreative II.5 challenge overview

Krallinger, Martin, Florian Leitner, and Alfonso Valencia

Proceedings of the BioCreative II.5 Challenge Evaluation Workshop, 2009

BioCreative II.5: Evaluation and ensemble system performance

Mardis, Scott, Florian Leitner, and Lynette Hirschman

Proceedings of the BioCreative II.5 Challenge Evaluation Workshop, 2009

SOA-Based Integration of Text Mining Services

Starlinger, Johann, Florian Leitner, Alfonso Valencia, and Ulf Leser

Proceedings of the 2009 World Conference on Services, 2009

Assessment of the second BioCreative PPI task: automatic extraction of protein-protein interactions

Krallinger, Martin, Florian Leitner, and Alfonso Valencia

Proceedings of the second BioCreative Challenge Evaluation Workshop, 2007

Exploiting text mining of biological annotations to extract protein interactions for the Semantic web

Krallinger, Martin, Carlos Rodriguez-Penagos, Florian Leitner, and Alfonso Valencia

Seventh International Workshop on Network Tools and Applications in Biology, 2007