

# 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

English: master

Spanish: proficient

## computing

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

data engineering:  
PostgreSQL, HDFS,  
Cassandra

language processing:  
StanfordNLP, Factorie,  
LingPipe, etc.

distributed computing:  
UIMA, Spark,  
ElasticSearch

software development:  
JetBrains IDEs,  
Jenkins/Travis,  
Gradle/Maven, 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  | <b>Leitner Catalytics, S.L.</b><br>Independent consultant (Data mining & NLP)<br>Delivering tailored, cloud-based data mining and natural language processing solutions to clients world-wide (in Java/Scala, Python, and R).  | Madrid, Spain       |
| 2014–2015 | <b>Technical University of Madrid (UPM)</b><br>Juan de la Cierva research fellow (Inference engine & NLP development)<br>Information extraction for neurobiology with Factorie, Spark, SciKit-Learn, OrientDB, and R. Statistical analyses (R). Teaching (R, Python, GATE).  | Madrid, Spain       |
| 2013–2014 | <b>National Cancer Research Centre (CNIO)</b><br>Post-doc (Data mining & NLP software development)<br>Extraction of the mammalian transcriptome with UIMA, Lucene, C/C++, SciKit-Learn, and D3.js.   | Madrid, Spain       |
| 2006–2013 | <b>National Cancer Research Centre (CNIO)</b><br>PhD student (Bioinformatics software development)<br>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 | <b>Institute for Molecular Pathology (IMP)</b><br>MSc student (Database administration and software/web development)<br>Data acquisition from biological resources (Perl, LISP); Web visualization (Struts) of those resources. Indexing system and query interface (Lucene).  | Vienna, Austria     |
| 2003–2003 | <b>CSO Lanifex GmbH</b><br>Consultant (Linux kernel and Firefox browser customization, web dev.)<br>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 | <b>European Media Labs (EML)</b><br>Python/Java application developer<br>3D visualization of protein structure annotations (Java). Structural biology research support using Python.   | Heidelberg, Germany |
| 2001–2002 | <b>Novartis Research Institute</b><br>Web developer and database administration<br>Perl-based bioinformatics framework development for a large Alpha cluster.  | Vienna, Austria     |
| 1997–2000 | <b>Pro-Design Paragliders</b><br>Global sales manager<br>Overseas production and intl. sales. PR, advertisements, sports events (product intros). In-house stock management system (Visual Basic).   | Innsbruck, Austria  |

## education

2012	<b>PhD Bioinformatics, with distinction</b> Online assessment of protein interaction information extraction systems.	Autonomous University of Madrid (UAM), Spain
2006	<b>MSc Molecular Biology, with distinction</b> Integration of biological databases and mining for post-translational modifications.	Vienna University, Austria
1995	<b>Advertisement &amp; Design School</b> Werbe- und Design Akademie	Innsbruck, Austria
1994	<b>Business High-school</b> Handelsschule Innsbruck	Innsbruck, Austria

## accomplishments

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

## communication skills

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

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

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

The CHEMDNER corpus of chemicals and drugs and its annotation principles

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

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

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

**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 Lægreid, 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

**BioCreative II.5: Evaluation and ensemble system performance**

Scott Mardis, Florian Leitner, Lynette Hirschman  
*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