Curriculum Vitae (CV)

Dr Philipp Mayr-Schlegel

Affiliation:
GESIS – Leibniz Institute for the Social Sciences
Unter Sachsenhausen 6-8
50667 Cologne, Germany
+49 - (0)221 - 47694 - 533
philipp.mayr@gesis.org

Born in Munich, Germany
Married, 5 children, residing in Wachtberg, Germany
Nationality: German

Philipp Mayr is a team leader (Information & Data Retrieval) at the GESIS department Knowledge Technologies for the Social Sciences (WTS). In the winter semester 2020/2021, he administrated the W3 professorship "Scientific Information Analytics" at the University of Göttingen, Institute of Computer Science. He has been a visiting professor for knowledge representation at University of Applied Sciences in Darmstadt, Department of Information Science & Engineering during 2009-2011. Philipp Mayr received his PhD in applied informetrics and information retrieval from the Berlin School of Library and Information Science at Humboldt University Berlin in 2009. To date, he has been awarded substantial research funding (PI, Co-PI) from national and European funding agencies. Philipp Mayr was General Chair of the 41st European Conference on Information Retrieval (ECIR 2019) in April 2019. He has published in top conferences and prestigious journals in the areas informetrics, information retrieval and digital libraries. His research group "Information and Data Retrieval" is working on methods and techniques of interactive information and dataset retrieval and maintains and further develops information systems and research infrastructures for the social sciences.

His research interests include: interactive IR, academic search, literature-based discovery, research data infrastructures, scholarly communication, non-textual ranking, bibliometric and scientometric methods, applied informetrics, science models in digital libraries, knowledge representation, semantic technologies, user studies, information behavior.

EDUCATION

February 2009	Graduation: Dr. phil. from Humboldt-Universität zu Berlin, Berlin School of Library and Information Science. Doctoral Thesis: "Re-Ranking auf Basis von Bradfordizing für die verteilte Suche in Digitalen Bibliotheken", Supervisor: Prof. Dr. Walther Umstätter (Humboldt-Universität zu Berlin) Co-Referent: Prof. Dr. Jürgen Krause (GESIS and Universität Koblenz) Grade "Magna cum laude"
2004-2008	Doctoral studies in the research group of Jürgen Krause "Modeling and Treatment of Semantic Heterogeneity" at GESIS - Leibniz Institute for the Social Sciences
March 2004	Graduation: Magister Artium (M.A.), Humboldt-Universität zu Berlin Grade 1.4
1997-2004	Studies of Library and Information Science, Computer Science and Social Sciences at Humboldt-Universität zu Berlin, Germany

1995-1996	Mandatory civil service at Asklepios Nordseeklinik Westerland/Sylt
1995	Graduation: Abitur at Gymnasium LSH Schloss Ising, Germany

SCIENTIFIC CAREER

Since 07.2010	Team leader at GESIS, leading the team "Information & Data Retrieval" with seven computer scientists (PostDocs and doctoral students)
10.2020-03.2021	Administration of the W3 professorship "Scientific Information Analytics" at the University of Göttingen, Institute of Computer Science
02.2011-12.2018	Deputy department head at GESIS, department "Knowledge technologies for the Social Sciences"
2012-2015	Senior lecturer at University of Applied Sciences Cologne (TH Köln), Faculty 03 of Information Science
2012	Offer of a full professorship (W2) in Information Science at the University of Applied Sciences Darmstadt (rejected)
10.2009-08.2011	Visiting professor for knowledge representation at Hochschule Darmstadt, Germany
2009-2010	Postdoctoral Researcher at GESIS, Germany
11.2004-01.2009	Research associate in the R&D department at GESIS, Germany
04.2001-10.2004	IT consultant and developer at HiSolutions AG, Berlin

RESEARCH STAYS

2017	Banaras Hindu University (India), Department of Computer Science (2 weeks)
2018	University of Toulouse (FR), Computer Science Department, IRIT (2 weeks)
2018	Banaras Hindu University (India), Department of Computer Science (2 weeks)

SCIENTIFIC PROJECTS AND GRANTS (SELECTION)

Grant budget for GESIS.

ConDATA <http://bit.ly/Condata>
Since 2019: DFG project "ConDATA" - Establishing Contextual Dataset Retrieval - transferring concepts from document to dataset retrieval. Objective: Employ and evaluate contextualization approaches (e.g. document similarity, session contexts) in the data retrieval infrastructure. Role: PI, 350,000 EUR.

OASE <http://bit.ly/OASE-project>
Since 2018: BMBF project "OASE" - "Open Access-Effekte" - Strukturelle und autorspezifische Einflussfaktoren auf den Impact von OA-Publikationen. Objective: Analyze the bibliometric and altmetric effects of scholarly publications being available Open Access. Bibliometric research with Web of Science, Scopus and Unpaywall data. Role: PI, together with Isabella Peters, ZBW Leibniz-Informationszentrum Wirtschaft & CAU Christian-Albrechts-Universität zu Kiel, 224,000 EUR.

POLLUX <http://bit.ly/Pollux-project>
Since 2016: DFG project "POLLUX" - Fachinformationsdienst Politikwissenschaft. Objective: Building a state-of-the-art literature retrieval portal for Political Science researchers. Role: PI, together with Maria Müller, Staats- und Universitätsbibliothek Bremen (SUUB), 332,000 EUR.

Competence Centre for Bibliometrics <http://bibliometrie.info/>
Since 2014: Project "Competence Centre for Bibliometrics". Objective: Provide, maintain and further develop a large-scale data infrastructure to facilitate bibliometric research. Role: member of the consortium, GESIS PI, together with DZHW, Fraunhofer ISI, Universität Bielefeld, FIZ Karlsruhe, Max-Planck Digital Library, ZB Jülich. Various funded projects:
- Scalable Disambiguation of Institutions for Web of Science, 23,000 EUR. <http://bit.ly/Institutions-project> Data release in preparation.
- Efficient Retrieval of Web of Science data with Elasticsearch, 17,000 EUR. <http://bit.ly/Elasticsearch-project>
- Analyzing Self-citations in Web of Science, 16,000 EUR. <http://bit.ly/Selfcitations-project>
- Vergabe von eindeutigen Autoren-IDs in WoS, 19,000 EUR. Open data released: Open Testbed for Author Name Disambiguation Evaluation http://dx.doi.org/10.7802/1234

EXCITE <http://bit.ly/EXCITE-project>
2016-2019: DFG project "EXCITE" - Extraction of Citations from PDF Documents. Objective: Development of a scalable toolchain of citation extraction and matching software with particular focus on the German-language social sciences. Role: PI, together with Steffen Staab, WeST – Institute for Web Science and Technologies, University of Koblenz-Landau, 305,000 EUR. Open data released for the OpenCitations project: https://github.com/exciteproject/EXpublisher

OSCOSS <http://bit.ly/OSCOSS-project>
2015-2019: DFG project "OSCOSS" - Opening Scholarly Communications in Social Sciences. Objective: Further development and integration of the collaborative document editor Fidus Writer with the Open Journal Systems and GESIS infrastructure. Role: PI, together with Sören Auer and Christoph Lange, University of Bonn, 171,000 EUR.

IRIS <https://sites.google.com/site/iris2019info/>
2019: Indo-German Workshop on Information Retrieval, Informetrics & Scientometrics (IRIS) funded by Indo-German Science & Technology Centre (IGSTC). Role: Chair, together with Vivek Kumar Singh, Banaras Hindu University, Varanasi, India, 20,000 EUR.

Scientotext <https://github.com/Scientotext/sciento-text>
2017-2019: Design of a Sciento-text Computational Framework for Retrieval and Contextual Recommendations of High Quality Scholarly Articles. Indo-German DST - DAAD Project-based Personnel Exchange Programme. Objective: Facilitating research collaboration and exchange stays of Indian and German researchers. Role: PI, together with Vivek Kumar Singh, Banaras Hindu University, Varanasi, India, 10,000 EUR. Open data released: https://github.com/Scientotext/PLOS-ONE-Dataset

AMUR <http://bit.ly/AMUR-project>
2015-2018: DFG project "AMUR" - User Guidance and System Tuning for Search Sessions. Objective: Implementation and evaluation of interactive search session support services in the productive scholarly digital library sowiport, Role: PI, together with Norbert Fuhr, Universität Duisburg-Essen, 168,000 EUR. Open data released: Sowiport User Search Sessions Data Set (SUSS), Sowiport user queries sample (SQS)

Linked.SWISSBIB <http://bit.ly/Swissbib>
2015-2017: Swiss SUC P-2 program project linked.swissbib.ch. Objective: Automatic interlinking of large-scale library catalog data and its enrichment with additional information (DBpedia and Virtual International Authority File - VIAF). Role: PI, together with René Schneider, Haute école de gestion de Genève, 111,000 EUR. Open data released.

IRM I and **II** <http://bit.ly/IRM-project>
From 2009-2013: DFG project "Value-Added Services for Information Retrieval" Part I and II. Objective: Implementation and evaluation of bibliometric-enhanced retrieval services. Role: PI and key scientist, 530,000 EUR.

Monitoring Bildungsforschung
From 2011-2014: Project „Entwicklung und Veränderungsdynamik eines heterogenen sozialwissenschaftlichen Feldes am Beispiel der Bildungsforschung", Role: PI, together with DIPF, IFQ, ZPID, 212,000 EUR. Open data released: https://git.gesis.org/open-data/solis-sofis

KoMoHe <http://bit.ly/KoMoHe-project>
From 2004-2008: BMBF project „Kompetenzzentrum Modellbildung und Heterogenitätsbehandlung". Role: key scientist. Open data released for the Linked Open Data Cloud https://lod-cloud.net/dataset/gesis-thesoz

SELECTED SERVICES/INFRASTRUCTURES

My team "Information & Data Retrieval" at GESIS and I maintain and further develop the following information systems and research data infrastructures:

GESIS search engine <https://search.gesis.org/>
Since 2017; new GESIS search engine integrating data sets with publications and other information like instruments, variables and questions from surveys (Technology: ElasticSearch)

ExploreData (beta)
Search demonstrator which offers an integrated search in variables and question of surveys in GESIS data catalogs

Pollux <https://www.pollux-fid.de/>
Since 2018, literature portal for Political Science researchers, together with SUUB Bremen

SSOAR full text repository <http://www.ssoar.info/>
2011-2019, Role: technical lead, project development, specification of new services and features, coordination with department WT@GESIS

sowiport search engine for the social sciences <discontinued>
2010-2018, Role: technical lead, project development, specification of new services and features coordination with department CSS@GESIS

LeibnizOpen full text repository <http://www.leibnizopen.de/home/>
2013-2018, Role: technical lead, project development, specification of new services and features coordination with department CSS@GESIS

ZIS Zusammenstellung sozialwissenschaftlicher Items und Skalen <http://zis.gesis.org/>
2016-2018, Role: coordination with department SDM@GESIS

SELECTED PEER-REVIEWED PUBLICATIONS

See list of publications in a separate list. See also https://philippmayr.github.io/#pub or http://scholar.google.com/citations?hl=en&user=RIB8suEAAAAJ

Kacem, A., Flatt, J. W., & Mayr, P. (2020). Tracking self-citations in academic publishing. Scientometrics. https://doi.org/10.1007/s11192-020-03413-9

Fraser, N., Momeni, F., Mayr, P., & Peters, I. (2020). The relationship between bioRxiv preprints, citations and altmetrics. Quantitative Science Studies. https://doi.org/10.1162/qss_a_00043

Zeng, M. L., & Mayr, P. (2019). Knowledge Organization Systems (KOS) in the Semantic Web: A Multi-Dimensional Review. International Journal on Digital Libraries, 20(3), 209–230. https://doi.org/10.1007/s00799-018-0241-2

Otto, W., Ghavimi, B., Mayr, P., Piryani, R., & Singh, V. K. (2019). Highly cited references in PLOS ONE and their in-text usage over time. Proceedings of the 17th International Conference on Scientometrics & Informetrics (ISSI 2019), 1531–1536. https://arxiv.org/abs/1903.11693

Kacem, A., & Mayr, P. (2018). Analysis of Search Stratagem Utilisation. Scientometrics, 116(2), 1383–1400. https://doi.org/10.1007/s11192-018-2821-8

Carevic, Z., Schüller, S., Mayr, P., & Fuhr, N. (2018). Contextualised Browsing in a Digital Library's Living Lab. In Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries (pp. 89–98). Fort Worth, Texas, USA: ACM New York, NY, USA. https://doi.org/10.1145/3197026.3197054

Bensmann, F., Zapilko, B., & Mayr, P. (2017). Interlinking Large-scale Library Data with Authority Records. Frontiers in Digital Humanities, 4. http://doi.org/10.3389/fdigh.2017.00005

Mutschke, P., & Mayr, P. (2015). Science Models for Search. A Study on Combining Scholarly Information Retrieval and Scientometrics. Scientometrics, 102(3), 2323–2345. http://doi.org/10.1007/s11192-014-1485-2

Mayr, P. (2013). Relevance distributions across Bradford Zones: Can Bradfordizing improve search? In J. Gorraiz, E. Schiebel, C. Gumpenberger, M. Hörlesberger, & H. Moed (Eds.), 14th International Society of Scientometrics and Informetrics Conference (pp. 1493–1505). Vienna, Austria. http://arxiv.org/abs/1305.0357

Mutschke, P., Mayr, P., Schaer, P., & Sure, Y. (2011). Science models as value-added services for scholarly information systems. Scientometrics, 89(1), 349–364. http://doi.org/10.1007/s11192-011-0430-x

Mayr, P., & Petras, V. (2008). Cross-concordances: terminology mapping and its effectiveness for information retrieval. In 74th IFLA World Library and Information Congress. Québec, Canada: IFLA. Retrieved from http://www.ifla.org/IV/ifla74/papers/129-Mayr_Petras-en.pdf

Mayr, P., & Walter, A.-K. (2007). An exploratory study of Google Scholar. Online Information Review, 31(6), 814–830. http://doi.org/10.1108/14684520710841784

OPEN DATASETS

PLOS ONE Citation Contexts (2019) https://github.com/Scientotext/PLOS-ONE-Dataset

Kacem, A., & Mayr, P. (2018). Sowiport User Search Sessions Data Set (SUSS) 16-17. https://git.gesis.org/amur/SUSS-16-17

Mayr, P. (2018). GESIS-SOLIS, GESIS-SOFIS databases. https://git.gesis.org/open-data/solis-sofis

Beel, J., Dinesh, S., Mayr, P., Carevic, Z., & Raghvendra, J. (2017). Stereotype and Most-Popular Recommendations in the Digital Library Sowiport [Dataset]. Harvard Dataverse. https://doi.org/10.7910/DVN/HFIV1A

Mayr, P. (2016a). Sowiport user queries sample (SQS). GESIS, Datorium. http://dx.doi.org/10.7802/1372

Mayr, P. (2016b). Sowiport User Search Sessions Data Set (SUSS). GESIS, Datorium. http://dx.doi.org/10.7802/1380

Momeni, F., & Mayr, P. (2016). An Open Testbed for Author Name Disambiguation Evaluation. GESIS, Datorium. http://dx.doi.org/10.7802/1234

Kaczmirek, L., & Mayr, P. (2015). German Bundestag Elections 2013: Twitter usage by electoral candidates. GESIS Data Archive. http://dx.doi.org/10.4232/1.12319

TheSoz Thesaurus for the Social Sciences (GESIS). (2013). https://lod-cloud.net/dataset/gesis-thesoz

EDITORIAL WORK AND REVIEWING

Guest Editor of the following special issues
- "Mining Scientific Papers Volume II: Knowledge Discovery and Data Exploitation" 2020 in **Frontiers in Research Metrics and Analytics**. https://www.frontiersin.org/research-topics/13388/mining-scientific-papers-volume-ii-knowledge-discovery-and-data-exploitation
- "Scholarly literature mining with Information Retrieval and Natural Language Processing" 2020 in **Scientometrics**. https://sites.google.com/view/scientometrics-si2019-bir
- "Mining Scientific Papers: NLP-enhanced Bibliometrics" 2019 in **Frontiers in Research Metrics and Analytics**. https://www.frontiersin.org/research-topics/7043/mining-scientific-papers-nlp-enhanced-bibliometrics
- "Bibliometric-enhanced information retrieval" 2018 in **Scientometrics**. https://link.springer.com/article/10.1007/s11192-018-2861-0
- "Bibliometric-Enhanced Information Retrieval and Natural Language Processing for Digital Libraries" 2018 in **International Journal on Digital Libraries**. https://link.springer.com/article/10.1007/s00799-017-0230-x
- "Knowledge Maps and Information Retrieval" 2017 in **International Journal on Digital Libraries** http://link.springer.com/10.1007/s00799-016-0204-4.
- "Networked Knowledge Organization Systems (NKOS)" 2016 in **International Journal on Digital Libraries** http://link.springer.com/10.1007/s00799-015-0167-x
- "Combining bibliometrics and information retrieval" 2015 in **Scientometrics** http://link.springer.com/article/10.1007/s11192-014-1484-3

I am a member of the editorial board of the journals
- Scientometrics,
- New Review of Hypermedia and Multimedia,
- International Journal on Digital Libraries,
- Frontiers in Research Metrics and Analytics,
- Information - Wissenschaft & Praxis.

I frequently review for the following list of:
- **International conferences**: JCDL, ECIR, SIGIR, TPDL, ISSI, ASIST, Open Science Conference, Dublin Core, ESWC, ISWC, LWDA, International Symposium of Information Science, iConference, WebSci
- **Journals**: Scientometrics, JASIST, Journal of Informetrics, PLOS ONE, IJDL, Information Processing and Management, Aslib Journal of Information Management, Knowledge and Information Systems

SELECTION OF WORKSHOPS

Chandrasekaran, M. K., Feigenblat, G., Freitag, D., Ghosal, T., Herrmannova, D., Hovy, E., … de Waard, A. (2020). First Workshop on Scholarly Document Processing (SDP 2020). Conference on Empirical Methods in Natural Language Processing (EMNLP 2020). **EMNLP 2020**, Role: Main organizer https://ornlcda.github.io/SDProc/

Zhang, C., Mayr, P., Lu, W., & Zhang, Y. (2020). Workshop on Extraction and Evaluation of Knowledge Entities from Scientific Documents (EEKE2020). **JCDL 2020**, Role: Main organizer https://eeke2020.github.io/

Cabanac, G., Frommholz, I., & Mayr, P. (2020). Bibliometric-Enhanced Information Retrieval 10th Anniversary Workshop Edition. **ECIR 2020**, Role: Main organizer https://sites.google.com/view/bir-ws/bir-2020

Chandrasekaran, M. et al. (2019). Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL 2019). Paris, France, **SIGIR 2019**, Role: Main organizer

Koesten, L. et al. (2018). DATA:SEARCH'18 - Searching Data on the Web. Ann Arbor, USA, **SIGIR 2018**, Role: Main organizer

Tudhope, D. et al. (2018) 18th Networked Knowledge Organisation Systems and Services (NKOS). Thessaloniki, Greece. **TPDL 2018**, Role: Main organizer

Cabanac, G. et al. (2016). Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL 2016). Newark, USA, **JCDL 2016**, Role: Main organizer

Atanassova, I., Bertin, M., & Mayr, P. (2015). First Workshop on Mining Scientific Papers: Computational Linguistics and Bibliometrics. Istanbul, Turkey. **ISSI 2015**, Role: Main organizer

Clarke, S. D., Greenberg, J., Lin, X., Mayr, P., & Tudhope, D. (2014). 13th Networked Knowledge Organisation Systems and Services (NKOS). London: ACM. **DL 2014**, Role: Main organizer

Mutschke, P., Scharnhorst, A., Guéret, C., Mayr, P., Hansen, P., & Slavic, A. (2014). Knowledge Maps and Information Retrieval (KMIR). London: ACM. **DL 2014**, Role: Co-organizer

Mayr, P., Scharnhorst, A., Larsen, B., Schaer, P., & Mutschke, P. (2014). First Bibliometric-enhanced Information Retrieval. Amsterdam. **ECIR 2014**, Role: Main organizer

Koch, T., Mayr, P., & Tudhope, D. (2011). 10th European Networked Knowledge Organization Systems (NKOS) Workshop. **TPDL 2011**, Role: Co-organizer

TEACHING AND SUPERVISION

Selected lectures, courses and seminars. See all teaching activities in a separate list.
- **Informetrics**, University of Göttingen, bachelor & master course, WS 2020/21
- **Digital Repositories**, University of Göttingen, bachelor & master course, WS 2020/21
- Information Retrieval, Fachhochschule Köln, bachelor course
- Database Retrieval / Web Retrieval, Fachhochschule Köln, bachelor course
- Information Seeking Behavior, Hochschule Darmstadt, master course and Fachhochschule Köln, bachelor course
- Evaluation of Retrieval Services: an introduction, Hochschule Darmstadt, master and bachelor course
- Thesauri and Ontologies, Hochschule Darmstadt, bachelor course with exercises
- Social Web and Bibliometry, Universität Koblenz, Summer Academy with exercises
- Digital Library Research, Hochschule Darmstadt, bachelor course with exercises
- Databases, Hochschule Darmstadt, bachelor course with exercises

Supervision of PhD thesis
- Philipp Schaer at GESIS, defended in May 2013
- Daniel Hienert at GESIS, defended in October 2013
- Zeljko Carevic at GESIS, defended in July 2020
- Fakhri Momeni at GESIS, ongoing

Supervision of Bachelor and Master thesis
- at University of Bonn, Computer Science/EIS
- at Fachhochschule Köln, department information science
- at Hochschule Darmstadt, department information science and engineering
- at Hochschule Bonn-Rhein-Sieg, department computer science

AWARDS

VFI-Förderungspreis 2009: Award for an excellent PhD thesis in information science