

RESEARCH CENTRE

**Inria Centre at Université Côte
d'Azur**

IN PARTNERSHIP WITH:

CNRS, Université Côte d'Azur

2024

ACTIVITY REPORT

Project-Team

WIMMICS

**Web-Instrumented Man-Machine
Interactions, Communities and Semantics**

IN COLLABORATION WITH: Laboratoire informatique, signaux systèmes
de Sophia Antipolis (I3S)

DOMAIN

Perception, Cognition and Interaction

THEME

**Data and Knowledge Representation and
Processing**

Inria

Contents

Project-Team WIMMICS	1
1 Team members, visitors, external collaborators	4
2 Overall objectives	7
2.1 Context and Objectives	7
2.2 Research Topics	7
3 Research program	8
3.1 Users Modeling and Designing Interaction on the Web and with AI systems	8
3.2 Communities and Social Media Interactions and Content Analysis on the Web and Linked Data	8
3.3 Vocabularies, Semantic Web and Linked Data Based Knowledge Extraction and Representation of Knowledge Graphs on the Web	9
3.4 Artificial Intelligence Processing: Learning, Analyzing and Reasoning on Heterogeneous Knowledge Graphs	9
4 Application domains	10
4.1 Social Semantic Web	10
4.2 Linked Data on the Web and on Intranets	10
4.3 Assisting Web-based Epistemic Communities	10
4.4 Linked Data for a Web of Diversity	11
4.5 Artificial Web Intelligence	11
4.6 Human-Data Interaction (HDI) on the Web	11
4.7 Web-augmented interactions with the world	12
4.8 Analysis of scientific co-authorship	12
5 Social and environmental responsibility	13
5.1 Footprint of research activities	13
5.2 Impact of research results	13
6 Highlights of the year	13
6.1 Awards	13
7 New software, platforms, open data	13
7.1 New software	13
7.1.1 ACTA	13
7.1.2 ARViz	14
7.1.3 Attune	14
7.1.4 CORESE-Core	14
7.1.5 CORESE-GUI	15
7.1.6 CORESE-Server	15
7.1.7 CORESE-Command	16
7.1.8 CROBORA	16
7.1.9 Datalens	16
7.1.10 DBpedia	17
7.1.11 Fuzzy labelling argumentation module	17
7.1.12 GUsT-3D	17
7.1.13 IndeGx	18
7.1.14 ISSA-pipeline	18
7.1.15 ISSA Visualization Web Application	19
7.1.16 KartoGraphI	19
7.1.17 Licentia	19
7.1.18 Metadatamatic	20
7.1.19 MGExplorer	20

7.1.20	Morph-xR2RML	21
7.1.21	Muvin	21
7.1.22	Olivaw	22
7.1.23	PEACE	22
7.1.24	plic2owl	23
7.1.25	RDFminer	23
7.1.26	SciLEX	23
7.1.27	SPARQL Micro-services	24
7.1.28	wam-studio	24
7.1.29	WebAudio tube guitar amp sims CLEAN, DISTO and METAL MACHINES	24
7.1.30	WheatGenomicsSLKG Visualization Web Application	25
7.1.31	Zoomathia KG Pipeline	25
7.2	Open data	25
8	New results	28
8.1	User Modeling and Designing Interaction	28
8.1.1	Interaction with extended reality	28
8.1.2	Visual exploration of datasets and models for AI	29
8.1.3	Interactive WebAudio applications	29
8.1.4	KartoGraphI	29
8.2	WisNote	30
8.3	Communities and Social Interactions Analysis	30
8.3.1	Online Hate Detection in Conversational Data	30
8.3.2	Unveiling the Hate: Generating Faithful and Plausible Explanations for Implicit and Subtle Hate Speech Detection	31
8.3.3	PEACE: Providing Explanations and Analysis for Combating Hate Expressions	31
8.3.4	Is Safer Better? The Impact of Guardrails on the Argumentative Strength of LLMs in Hate Speech Countering	31
8.3.5	Argumentation Quality Assessment	32
8.3.6	DISPUTool 2.0: A Modular Architecture for Multi-Layer Argumentative Analysis of Political Debates	32
8.3.7	Argument-based Detection and Classification of Fallacies in Political Debates	32
8.3.8	FALCON: A multi-label graph-based dataset for fallacy classification in the COVID-19 infodemic	33
8.3.9	Augmenting participation, co-creation, trust and transparency in Deliberative Democracy at all scales through Argument Mining	33
8.3.10	Controversy and influence in the Ukraine war: a study of argumentation and counter-argumentation through Artificial Intelligence	34
8.3.11	Argument-structured Justification Generation for Explainable Fact-checking	34
8.3.12	Study of the Attention Economy, its detrimental impacts and leads for regulation	34
8.3.13	CasiMedicos-Arg: A Medical Question Answering Dataset Annotated with Explanatory Argumentative Structures	35
8.4	Vocabularies, Semantic Web and Linked Data Based Knowledge Representation and Artificial Intelligence Formalisms on the Web	35
8.4.1	Semantic Web for Life Sciences	35
8.4.2	W3C Data activity and AC Rep	35
8.4.3	AutomaZoo: Automatic annotation of an ancient Zoological Corpus	36
8.4.4	A Unified approach to publish semantic annotations of agricultural documents as knowledge graphs	36
8.4.5	Semantic Web for Biodiversity	37
8.4.6	Ontology engineering: tooling and methodology	37
8.5	Analyzing and Reasoning on Heterogeneous Semantic Graphs	37
8.5.1	Corese Semantic Web Factory	37
8.5.2	Explanatory Argumentation in Natural Language for Correct and Incorrect Medical Diagnoses	38

8.5.3	ACTA Module Upgrade	38
8.5.4	RDF Mining	38
8.5.5	Capturing Geospatial Knowledge from Real-Estate Advertisements	39
8.5.6	Outlier Detection in MET data using Subspace Outlier Detection Method	39
8.5.7	Estimating User's Knowledge Gain in Search-as-Learning Using Knowledge Graphs	40
8.5.8	ISSA: semantic indexing of scientific articles and advanced services	40
8.5.9	IndeGx	40
8.5.10	Learning Pattern-Based Extractors from Natural Language and Knowledge Graphs: Applying Large Language Models to Wikipedia and Linked Open Data	41
8.5.11	AI agent to convert natural language questions into SPARQL queries	41
8.5.12	Hybridizing machine learning and knowledge graphs: injection of relation signatures	42
8.5.13	Enriching benchmark datasets by generating synthetic ontologies and knowledge graphs	42
8.5.14	Reusing Wikidata to build new knowledge graphs: the KGPrune platform	43
8.5.15	Formal Argumentation: How can we understand and deal with implicit arguments?	43
8.5.16	How to reason with probabilistic and temporal information?	43
9	Bilateral contracts and grants with industry	44
9.1	Bilateral contracts with industry	44
10	Partnerships and cooperations	46
10.1	International research visitors	46
10.1.1	Visits of international scientists	46
10.1.2	Visits to international teams	46
10.2	European initiatives	47
10.2.1	H2020 projects	47
10.2.2	Other european programs/initiatives	48
10.3	National initiatives	49
10.4	Regional initiatives	52
11	Dissemination	52
11.1	Promoting scientific activities	52
11.1.1	Scientific events: organization	52
11.1.2	Scientific events: selection	53
11.1.3	Journal	54
11.1.4	Invited talks	55
11.1.5	Scientific expertise	56
11.1.6	Research administration	56
11.1.7	Promoting Open Science practices	56
11.2	Teaching - Supervision - Juries	56
11.2.1	Teaching	56
11.2.2	Supervision	59
11.2.3	Juries	61
11.3	Popularization	62
11.3.1	Productions (articles, videos, podcasts, serious games, ...)	62
11.3.2	Participation in Live events	63
12	Scientific production	63
12.1	Major publications	63
12.2	Publications of the year	65
12.3	Cited publications	72

Project-Team WIMMICS

Creation of the Project-Team: 2013 July 01

Keywords

Computer sciences and digital sciences

- A1.2.9. – Social Networks
- A1.3.1. – Web
- A3.1.1. – Modeling, representation
- A3.1.2. – Data management, quering and storage
- A3.1.3. – Distributed data
- A3.1.4. – Uncertain data
- A3.1.6. – Query optimization
- A3.1.7. – Open data
- A3.1.10. – Heterogeneous data
- A3.1.11. – Structured data
- A3.2. – Knowledge
- A3.2.1. – Knowledge bases
- A3.2.2. – Knowledge extraction, cleaning
- A3.2.3. – Inference
- A3.2.4. – Semantic Web
- A3.2.5. – Ontologies
- A3.2.6. – Linked data
- A3.3.2. – Data mining
- A3.4. – Machine learning and statistics
- A3.4.1. – Supervised learning
- A3.4.6. – Neural networks
- A3.4.8. – Deep learning
- A3.5. – Social networks
- A3.5.1. – Analysis of large graphs
- A3.5.2. – Recommendation systems
- A5.1. – Human-Computer Interaction
- A5.1.1. – Engineering of interactive systems
- A5.1.2. – Evaluation of interactive systems
- A5.1.9. – User and perceptual studies
- A5.2. – Data visualization
- A5.7.2. – Music
- A5.8. – Natural language processing

- A7.1.3. – Graph algorithms
- A7.2.2. – Automated Theorem Proving
- A8.2.2. – Evolutionary algorithms
- A9.1. – Knowledge
- A9.2. – Machine learning
- A9.4. – Natural language processing
- A9.6. – Decision support
- A9.7. – AI algorithmics
- A9.8. – Reasoning
- A9.9. – Distributed AI, Multi-agent
- A9.10. – Hybrid approaches for AI

Other research topics and application domains

- B1.2.2. – Cognitive science
- B2. – Health
- B5.1. – Factory of the future
- B5.6. – Robotic systems
- B5.8. – Learning and training
- B6.3.1. – Web
- B6.3.2. – Network protocols
- B6.3.4. – Social Networks
- B6.4. – Internet of things
- B6.5. – Information systems
- B8.5. – Smart society
- B8.5.1. – Participative democracy
- B9. – Society and Knowledge
- B9.1. – Education
- B9.1.1. – E-learning, MOOC
- B9.1.2. – Serious games
- B9.2. – Art
- B9.2.1. – Music, sound
- B9.3. – Medias
- B9.5.1. – Computer science
- B9.5.6. – Data science
- B9.6. – Humanities
- B9.6.1. – Psychology
- B9.6.2. – Juridical science
- B9.6.5. – Sociology
- B9.6.7. – Geography
- B9.6.8. – Linguistics

B9.6.9. – Political sciences

B9.6.10. – Digital humanities

B9.7. – Knowledge dissemination

B9.7.1. – Open access

B9.7.2. – Open data

B9.9. – Ethics

B9.10. – Privacy

1 Team members, visitors, external collaborators

Research Scientists

- Fabien Gandon [Team leader, INRIA, Senior Researcher]
- Pierre-Antoine Champin [INRIA, Researcher]
- Victor David [INRIA, ISFP]
- Federica Granese [INRIA, Starting Research Position, from Dec 2024]
- Serena Villata [CNRS, Senior Researcher]

Faculty Members

- Michel Buffa [UNIV COTE D'AZUR, Professor]
- Elena Cabrio [UNIV COTE D'AZUR, Professor]
- Catherine Faron [UNIV COTE D'AZUR, Professor]
- Aline Menin [UNIV COTE D'AZUR, Associate Professor]
- Pierre Monnin [UNIV COTE D'AZUR, Associate Professor]
- Anaïs Ollagnier [UNIV COTE D'AZUR, Associate Professor]
- Andrea Tettamanzi [UNIV COTE D'AZUR, Professor]
- Marco Winckler [UNIV COTE D'AZUR, Professor, HDR]

Post-Doctoral Fellows

- Cristian Cardellino [UNIV COTE D'AZUR, until Jul 2024]
- Sofiane Elguendouze [CNRS, Post-Doctoral Fellow, from Sep 2024]
- Yingxue Fu [UNIV COTE D'AZUR]
- Yousouf Taghzouti [UNIV COTE D'AZUR, Post-Doctoral Fellow, from Nov 2024]

PhD Students

- Ali Ballout [UNIV COTE D'AZUR, until Jun 2024]
- Helena Bonaldi [UNIV TRENTE, until Jan 2024]
- Lucie Cadorel [SEPTEO, CIFRE, until Jun 2024]
- Rony Dupuy Charles [DORIANE, CIFRE, until Sep 2024]
- Greta Damo [UNIV COTE D'AZUR]
- Deborah Dore [CNRS, from Sep 2024]
- Remi Felin [UNIV COTE D'AZUR, until Nov 2024]
- Pierpaolo Goffredo [CNRS, from Sep 2024]
- Guillaume Meroue [INRIA, from Sep 2024]
- Cyprien Michel-Deletie [ENS DE LYON, from Sep 2024]

- Benjamin Molinet [UNIV COTE D'AZUR]
- Nicolas Ocampo [UNIV COTE D'AZUR]
- Elena Palmieri [UNIV BOLOGNA, from Jul 2024 until Oct 2024]
- Clement Quere [UNIV COTE D'AZUR]
- Shihong Ren [UNIV JEAN MONNET, until Apr 2024]
- Celian Ringwald [INRIA]
- Nicolas Robert [UNIV COTE D'AZUR, from Oct 2024]
- Nicolas Robert [INRIA, from Sep 2024 until Sep 2024]
- Ekaterina Sviridova [UNIV COTE D'AZUR]
- Maroua Tikat [UNIV COTE D'AZUR, until Sep 2024]
- Xiaou Wang [CNRS]

Technical Staff

- Arnaud Barbe [INRIA, UNIV COTE D'AZUR, Engineer]
- Anna Bobasheva [INRIA, Engineer]
- Remi Ceres [INRIA, Engineer]
- Mariana Eugenia Chaves Espinoza [CNRS, Engineer]
- Theo Alkibiades Collias [UNIV COTE D'AZUR, Engineer, until Jul 2024]
- Antoine De Smidt [CNRS, Engineer, until Jun 2024]
- Molka Dhouib [UNIV COTE D'AZUR, Engineer, from Apr 2024]
- Molka Dhouib [INRIA, Engineer, until Mar 2024]
- Florent Jaillet [I3S, Engineer, from Apr 2024]
- Maxime Lecoq [INRIA, Engineer, until Sep 2024]
- Pierre Maillot [INRIA, Engineer]
- Franck Michel [CNRS, Engineer]
- Nicolas Robert [INRIA, Engineer, until Jun 2024]

Interns and Apprentices

- Charafedinne Achir [I3S, Intern, from Jun 2024 until Aug 2024]
- Muhammad Ahmed [I3S, Intern, from Jun 2024 until Aug 2024]
- Hajar Bakarou [I3S, Intern, from May 2024 until Oct 2024]
- Hamza Belgroun [INRIA, Intern, from May 2024 until Aug 2024, SCIENCE PO]
- Hugo Carton [INRIA, Apprentice, until Sep 2024]
- Maëlle Debard [I3S, Intern, from Jun 2024 until Jul 2024]
- Samuel Demont [I3S, Intern, from May 2024 until Aug 2024]

- Deborah Dore [UNIV SAPIENZA , Intern, until Feb 2024]
- Mohamed Sinane El Messoussi [I3S, Intern, from May 2024 until Oct 2024]
- Quentin Escobar [UNIV COTE D'AZUR, Apprentice, from Sep 2024]
- Irina Gokhaeva [UNIV COTE D'AZUR, Intern, from Feb 2024 until Mar 2024]
- Erwan Hain [INRIA, Apprentice, from Mar 2024]
- Ayoub Hofr [I3S, Intern, from May 2024 until Oct 2024]
- Ekaterina Kolos [CNRS]
- Guillaume Meroue [UNIV COTE D'AZUR, Intern, from Mar 2024 until Jul 2024]
- Jeremy Moncada [INRIA, Intern, from Apr 2024 until Jul 2024]
- Dheeraj Parkash [I3S, Intern, from Jun 2024 until Aug 2024]
- Quentin Scordo [INRIA, Intern, from Apr 2024 until Sep 2024]
- Manuel Vimercati [UNIV MILANO, Intern]

Administrative Assistants

- Delphine Robache [INRIA]
- Lionel Tavanti [UNIV COTE D'AZUR, I3S]

Visiting Scientists

- Xabier Garmendia [UNIV BASQUE COUNTRY, from Apr 2024 until Jul 2024]
- Dario Malchiodi [UNIV MILANO, from Nov 2024]
- Victor Hugo Nascimento Rocha [UNIV SAO PAULO, from Mar 2024 until Aug 2024]

External Collaborators

- Hanna Abi Akl [DSTI]
- Andrei Ciortea [UNIV ST GALLEN, Assistant Professor]
- Olivier Corby [INRIA, Retired Researcher]
- Nicolas Delaforge [Probabl, from Sep 2024]
- Alain Giboin [INRIA, Retired Researcher]
- Freddy Lecue [JP MORGAN, AI Research Director]
- Christopher Leturc [UNIV COTE D'AZUR]
- Stefan Sarkadi [KINGS COLLEGE LONDON]

2 Overall objectives

2.1 Context and Objectives

The World Wide Web has transformed into a virtual realm where individuals and software interact in diverse communities. The Web has the potential to become the collaborative space for both natural and artificial intelligence, thereby posing the challenge of supporting these global interactions. The large-scale, mixed interactions inherent in this scenario present a plethora of issues that must be addressed through multidisciplinary approaches [115].

One particular problem is to reconcile the formal semantics of computer science (such as logics, ontologies, typing systems, protocols, etc.) on which the Web architecture is built, with the soft semantics of human interactions (such as posts, tags, status, relationships, etc.) that form the foundation of Web content. This requires a holistic approach that considers both the technical and social aspects of the Web, in order to ensure that the interactions between computational and natural intelligence are seamless and meaningful.

Wimmics proposes a range of models and methods to bridge the gap between formal semantics and social semantics on the World Wide Web [116], in order to address some of the challenges associated with constructing a universal space that connects various forms of intelligence.

From a formal modeling point of view, one of the consequences of the evolutions of the Web is that the initial graph of linked pages has been joined by a growing number of other graphs. This initial graph is now mixed with sociograms capturing the social network structure, workflows specifying the decision paths to be followed, browsing logs capturing the trails of our navigation, service compositions specifying distributed processing, open data linking distant datasets, etc. Moreover, these graphs are not available in a single central repository but distributed over many different sources. Some sub-graphs are small and local (e.g. a user's profile on a device), some are huge and hosted on clusters (e.g. Wikipedia), some are largely stable (e.g. thesaurus of Latin), some change several times per second (e.g. social network statuses), etc. Moreover, each type of network of the Web is not an isolated island. Networks interact with each other: the networks of communities influence the message flows, their subjects and types, the semantic links between terms interact with the links between sites and vice-versa, etc.

Not only do we need means to represent and analyze each kind of graphs, we also do need means to combine them and to perform multi-criteria analysis on their combination. Wimmics contributes to these challenges by: (1) proposing multidisciplinary approaches to analyze and model the many aspects of these intertwined information systems, their communities of users and their interactions; (2) formalizing and reasoning on these models using graphs-based knowledge representation from the semantic Web [1] to propose new analysis tools and indicators, and to support new functionalities and better management. In a nutshell, the first research direction looks at models of systems, users, communities and interactions while the second research direction considers formalisms and algorithms to represent them and reason on their representations.

2.2 Research Topics

The research objectives of Wimmics can be grouped according to four topics that we identify in reconciling social and formal semantics on the Web:

Topic 1 - users modeling and designing interaction on the Web and with AI systems: The general research question addressed by this objective is “*How do we improve our interactions with a semantic and social Web more and more complex and dense?*”. Wimmics focuses on specific sub-questions: “How can we capture and model the users’ characteristics?” “How can we represent and reason with the users’ profiles?” “How can we adapt the system behaviors as a result?” “How can we design new interaction means?” “How can we evaluate the quality of the interaction designed?”. This topic includes a long-term research direction in Wimmics on information visualization of semantic graphs on the Web. The general research question addressed in this last objective is “*How to represent the inner and complex relationships between data obtained from large and multivariate knowledge graph?*”. Wimmics focuses on several sub-questions: “Which visualization techniques are suitable (from a user point of view) to support the exploration and the analysis of large graphs?” “How to identify the new knowledge created by users during the exploration of knowledge graph?” “How to formally describe the dynamic transformations allowing

to convert raw data extracted from the Web into meaningful visual representations?” “How to guide the analysis of graphs that might contain data with diverse levels of accuracy, precision and interestingness to the users?”

Topic 2 - communities and social interactions and content analysis on the Web and Linked Data: The general question addressed in this second objective is “*How can we manage the collective activity on social media?*”. Wimmics focuses on the following sub-questions: “How do we analyze the social interaction practices and the structures in which these practices take place?” “How do we capture the social interactions and structures?” “How can we formalize the models of these social constructs?” “How can we analyze and reason on these models of the social activity?”

Topic 3 - vocabularies, semantic Web and linked data based knowledge extraction and representation with knowledge graphs on the Web: The general question addressed in this third objective is “*What are the needed schemas and extensions of the semantic Web formalisms for our models?*”. Wimmics focuses on several sub-questions: “What kinds of formalism are the best suited for the models of the previous section?” “What are the limitations and possible extensions of existing formalisms?” “What are the missing schemas, ontologies, vocabularies?” “What are the links and possible combinations between existing formalisms?” We also address the question of knowledge extraction and especially AI and NLP methods to extract knowledge from text. In a nutshell, an important part of this objective is to formalize the models identified in the previous objectives as typed graphs and to populate them in order for software to exploit these knowledge graphs in their processing (in the next objective).

Topic 4 - artificial intelligence processing: learning, analyzing and reasoning on heterogeneous knowledge graphs: The general research question addressed in this objective is “*What are the algorithms required to analyze and reason on the heterogeneous graphs we obtained?*”. Wimmics focuses on several sub-questions: “How do we analyze graphs of different types and their interactions?” “How do we support different graph life-cycles, calculations and characteristics in a coherent and understandable way?” “What kind of algorithms can support the different tasks of our users?”.

3 Research program

3.1 Users Modeling and Designing Interaction on the Web and with AI systems

Wimmics focuses on interactions of ordinary users with ontology-based knowledge systems, with a preference for semantic Web formalisms and Web 2.0 applications. We specialize in interaction design and evaluation methods to Web application tasks such as searching, browsing, contributing or protecting data. The team is especially interested in using semantics in assisting the interactions. We propose knowledge graph representations and algorithms to support interaction adaptation, for instance for context-awareness or intelligent interactions with machine. We propose and evaluate Web-based visualization techniques for linked data, querying, reasoning, explaining and justifying. Wimmics also integrates natural language processing approaches to support natural language based interactions. We rely on cognitive studies to build models of the system, the user and the interactions between users through the system, in order to support and improve these interactions. We extend the user modeling technique known as *Personas* where user models are represented as specific, individual humans. *Personas* are derived from significant behavior patterns (i.e., sets of behavioral variables) elicited from interviews with and observations of users (and sometimes customers) of the future product. Our user models specialize *Personas* approaches to include aspects appropriate to Web applications. Wimmics also extends user models to capture very different aspects (e.g. emotional states).

3.2 Communities and Social Media Interactions and Content Analysis on the Web and Linked Data

The domain of social network analysis is a whole research domain in itself and Wimmics targets what can be done with typed graphs, knowledge representations and social models. We also focus on the specificity of social Web and semantic Web applications and in bridging and combining the different social Web data structures and semantic Web formalisms. Beyond the individual user models, we rely on social studies to build models of the communities, their vocabularies, activities and protocols in order to identify where

and when formal semantics is useful. We propose models of collectives of users and of their collaborative functioning extending the collaboration personas and methods to assess the quality of coordination interactions and the quality of coordination artifacts. We extend and compare community detection algorithms to identify and label communities of interest with the topics they share. We propose mixed representations containing social semantic representations (e.g. folksonomies) and formal semantic representations (e.g. ontologies) and propose operations that allow us to couple them and exchange knowledge between them. Moving to social interaction we develop models and algorithms to mine and integrate different yet linked aspects of social media contributions (opinions, arguments and emotions) relying in particular on natural language processing and argumentation theory. To complement the study of communities we rely on multi-agent systems to simulate and study social behaviors. Finally we also rely on Web 2.0 principles to provide and evaluate social Web applications.

3.3 Vocabularies, Semantic Web and Linked Data Based Knowledge Extraction and Representation of Knowledge Graphs on the Web

For all the models we identified in the previous sections, we rely on and evaluate knowledge representation methodologies and theories, in particular ontology-based modeling. We also propose models and formalisms to capture and merge representations of different levels of semantics (e.g. formal ontologies and social folksonomies). The important point is to allow us to capture those structures precisely and flexibly and yet create as many links as possible between these different objects. We propose vocabularies and semantic Web formalizations for all the aspects that we model and we consider and study extensions of these formalisms when needed. The results have all in common to pursue the representation and publication of our models as linked data. We also contribute to the extraction, transformation and linking of existing resources (informal models, databases, texts, etc.) to publish knowledge graphs on the Semantic Web and as Linked Data. Examples of aspects we formalize include: user profiles, social relations, linguistic knowledge, bio-medical data, business processes, derivation rules, temporal descriptions, explanations, presentation conditions, access rights, uncertainty, emotional states, licenses, learning resources, etc. At a more conceptual level we also work on modeling the Web architecture with philosophical tools so as to give a realistic account of identity and reference and to better understand the whole context of our research and its conceptual cornerstones.

3.4 Artificial Intelligence Processing: Learning, Analyzing and Reasoning on Heterogeneous Knowledge Graphs

One of the characteristics of Wimmics is to rely on graph formalisms unified in an abstract graph model and operators unified in an abstract graph machine to formalize and process semantic Web data, Web resources, services metadata and social Web data. In particular Corese, the core software of Wimmics, maintains and implements that abstraction. We propose algorithms to process the mixed representations of the previous section. In particular we are interested in allowing cross-enrichment between them and in exploiting the life cycle and specificity of each one to foster the life-cycles of the others. Our results all have in common to pursue analyzing and reasoning on heterogeneous knowledge graphs issued from social and semantic Web applications. Many approaches emphasize the logical aspect of the problem especially because logics are close to computer languages. We defend that the graph nature of Linked Data on the Web and the large variety of types of links that compose them call for typed graphs models. We believe the relational dimension is of paramount importance in these representations and we propose to consider all these representations as fragments of a typed graph formalism directly built above the Semantic Web formalisms. Our choice of a graph based programming approach for the semantic and social Web and of a focus on one graph based formalism is also an efficient way to support interoperability, genericity, uniformity and reuse.

4 Application domains

4.1 Social Semantic Web

A number of evolutions have changed the face of information systems in the past decade but the advent of the Web is unquestionably a major one and it is here to stay. From an initial wide-spread perception of a public documentary system, the Web as an object turned into a social virtual space and, as a technology, grew as an application design paradigm (services, data formats, query languages, scripting, interfaces, reasoning, etc.). The universal deployment and support of its standards led the Web to take over nearly all of our information systems. As the Web continues to evolve, our information systems are evolving with it.

Today in organizations, not only almost every internal information system is a Web application, but these applications more and more often interact with external Web applications. The complexity and coupling of these Web-based information systems call for specification methods and engineering tools. From capturing the needs of users to deploying a usable solution, there are many steps involving computer science specialists and non-specialists.

We defend the idea of relying on Semantic Web formalisms to capture and reason on the models of these information systems supporting the design, evolution, interoperability and reuse of the models and their data as well as the workflows and the processing.

4.2 Linked Data on the Web and on Intranets

With billions of triples online (see Linked Open Data initiative), the Semantic Web is providing and linking open data at a growing pace and publishing and interlinking the semantics of their schemas. Information systems can now tap into and contribute to this Web of data, pulling and integrating data on demand. Many organizations also started to use this approach on their intranets leading to what is called linked enterprise data.

A first application domain for us is the publication and linking of data and their schemas through Web architectures. Our results provide software platforms to publish and query data and their schemas, to enrich these data in particular by reasoning on their schemas, to control their access and licenses, to assist the workflows that exploit them, to support the use of distributed datasets, to assist the browsing and visualization of data, etc.

Examples of collaboration and applied projects include: Corese, DBpedia.fr, DekaLog, D2KAB, MonaLIA.

4.3 Assisting Web-based Epistemic Communities

In parallel with linked open data on the Web, social Web applications also spread virally (e.g. Facebook growing toward 1.5 billion users) first giving the Web back its status of a social read-write media and then putting it back on track to its full potential of a virtual place where to act, react and interact. In addition, many organizations are now considering deploying social Web applications internally to foster community building, expert cartography, business intelligence, technological watch and knowledge sharing in general.

By reasoning on the Linked Data and the semantics of the schemas used to represent social structures and Web resources, we provide applications supporting communities of practice and interest and fostering their interactions in many different contexts (e-learning, business intelligence, technical watch, etc.).

We use typed graphs to capture and mix: social networks with the kinds of relationships and the descriptions of the persons; compositions of Web services with types of inputs and outputs; links between documents with their genre and topics; hierarchies of classes, thesauri, ontologies and folksonomies; recorded traces and suggested navigation courses; submitted queries and detected frequent patterns; timelines and workflows; etc.

Our results assist epistemic communities in their daily activities such as biologists and policymakers exchanging results, business intelligence and technological watch networks informing companies, engineers interacting on a project, conference attendees, students following the same course, tourists

visiting a region, mobile experts on the field, etc. Examples of collaboration and applied projects: ISSA, TeachOnMars, CREEP, ATTENTION, ORBIS, CIGAIA.

4.4 Linked Data for a Web of Diversity

We intend to build on our results on explanations (provenance, traceability, justifications) and to continue our work on opinions and arguments mining toward the global analysis of controversies and online debates. One result would be to provide new search results encompassing the diversity of viewpoints and providing indicators supporting opinion and decision making and ultimately a Web of trust. Trust indicators may require collaborations with teams specialized in data certification, cryptography, signature, security services and protocols, etc. This will raise the specific problem of interaction design for security and privacy. In addition, from the point of view of the content, this requires to foster the publication and coexistence of heterogeneous data with different points of views and conceptualizations of the world. We intend to pursue the extension of formalisms to allow different representations of the world to co-exist and be linked and we will pay special attention to the cultural domain and the digital humanities. Examples of collaboration and applied projects: ACTA, DISPUTOOL.

4.5 Artificial Web Intelligence

We intend to build on our experience in artificial intelligence (knowledge representation, reasoning) and distributed artificial intelligence (multi-agent systems - MAS) to enrich formalisms and propose alternative types of reasoning (graph-based operations, reasoning with uncertainty, inductive reasoning, non-monotonic, etc.) and alternative architectures for linked data with adequate changes and extensions required by the open nature of the Web. There is a clear renewed interest in AI for the Web in general and for Web intelligence in particular. Moreover, distributed AI and MAS provide both new architectures and new simulation platforms for the Web. At the macro level, the evolution accelerated with HTML5 toward Web pages as full applications and direct Page2Page communication between browser clearly is a new area for MAS and P2P architectures. Interesting scenarios include the support of a strong decentralization of the Web and its resilience to degraded technical conditions (downscaling the Web), allowing pages to connect in a decentralized way, forming a neutral space, and possibly going offline and online again in erratic ways. At the micro level, one can imagine the place **RDF (Resource Description Framework)** and **SPARQL (SPARQL Protocol and RDF Query Language)** could take as data model and programming model in the virtual machines of these new Web pages and, of course, in the Web servers. RDF is also used to serialize and encapsulate other languages and becomes a pivot language in linking very different applications and aspects of applications. Example of collaboration and applied projects: HyperAgents, DekaLog, AI4EU, AI4Media.

4.6 Human-Data Interaction (HDI) on the Web

We need more interaction design tools and methods for linked data access and contribution. We intend to extend our work on exploratory search coupling it with visual analytics to assist sense making. It could be a continuation of the Gephi extension that we built targeting more support for non experts to access and analyze data on a topic or an issue of their choice. More generally speaking SPARQL is inappropriate for common users and we need to support a larger variety of interaction means with linked data. We also believe linked data and natural language processing (NLP) have to be strongly integrated to support natural language based interactions. Linked Open Data (LOD) for NLP, NLP for LOD and Natural Dialog Processing for querying, extracting and asserting data on the Web is a priority to democratize its use. Micro accesses and micro contributions are important to ensure public participation and also call for customized interfaces and thus for methods and tools to generate these interfaces. In addition, the user profiles are being enriched now with new data about the user such as his/her current mental and physical state, the emotion he/she just expressed or his/her cognitive performances. Taking into account this information to improve the interactions, change the behavior of the system and adapt the interface is a promising direction. And these human-data interaction means should also be available for "small data", helping the user to manage his/her personal information and to link it to public or collective one, maintaining his/her personal and private perspective as a personal Web of data. Finally, the continuous

knowledge extractions, updates and flows add the additional problem of representing, storing, querying and interacting with dynamic data. Examples of collaboration and applied projects: WASABI, MuvIn, LDViz.

4.7 Web-augmented interactions with the world

The Web continues to augment our perception and interaction with reality. In particular, Linked Open Data enable new augmented reality applications by providing data sources on almost any topic. The current enthusiasm for the Web of Things, where every object has a corresponding Web resource, requires evolutions of our vision and use of the Web architecture. This vision requires new techniques as the ones mentioned above to support local search and contextual access to local resources but also new methods and tools to design Web-based human devices interactions, accessibility, etc. These new usages are placing new requirements on the Web Architecture in general and on the semantic Web models and algorithms in particular to handle new types of linked data. They should support implicit requests considering the user context as a permanent query. They should also simplify our interactions with devices around us jointly using our personal preferences and public common knowledge to focus the interaction on the vital minimum that cannot be derived in another way. For instance, the access to the Web of data for a robot can completely change the quality of the interactions it can offer. Again, these interactions and the data they require raise problems of security and privacy. Examples of collaboration and applied projects: ALOOF, AZKAR, MoreWAIS.

4.8 Analysis of scientific co-authorship

Over the last decades, scientific research has matured and diversified. In all areas of knowledge, we observe an increasing number of scientific publications, a rapid development of ever more specialized conferences and journals, and the creation of dynamic collaborative networks that cross borders and evolve over time. In this context, analyzing scientific publications and the resulting inner co-authorship networks is a major issue for the sustainability of scientific research. To illustrate this, let us consider what happens in the context of the COVID-19 pandemics, when the whole scientific community engaged numerous fields of research to contribute in a common effort to study, understand and fight the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). In order to support the scientific community, many datasets covering the publications about coronaviruses and related diseases have been compiled. In a short time, the number of publications available (over 200,000+ and still increasing) suggests that it is impossible for any researcher to examine every publication and extract the relevant information.

By reasoning on the Linked Data and Web semantic schemas, we investigate methods and tools to assist users on finding relevant publications to answer their research questions. Hereafter we present some example of typical domain questions and how we can contribute to the matter.

- How to find relevant publication in huge datasets? We investigate the use of association rules as a suitable solution to identify relevant scientific publications. By extracting association rules that determine the co-occurrence between terms in a text, it is possible to create clusters of scientific publications that follow a certain pattern; users can focus the search on clusters that contain the terms of interests rather than search the whole dataset.
- How to explain the contents of scientific publications? By reasoning on the Linked Data and Web semantic schemas, we investigate methods for the creation and exploration of argument graphs that describe association and development of ideas in scientific papers.
- How to understand the impact of co-authorship (collaboration of one or more authors) in the development of scientific knowledge? For that, we proposed visualization techniques that allows the description of co-authorship networks describing the clusters of collaborations that evolve over time. Co-authorship networks can inform both collaboration between authors and institutions.

Currently, the analysis of co-publications has been performed over two majors datasets: Hal open archive, the Covid-on-the-Web datasets, and the Agritrop (CIRAD's open dataset).

5 Social and environmental responsibility

5.1 Footprint of research activities

The team now integrates footprint metrics in its evaluations and comparisons of methods. For instance, in [76], we monitored the training time and the carbon cost for training a knowledge graph extraction model.

5.2 Impact of research results

We are especially interested in identifying Knowledge Graphs tasks for which SMLs (Small Language Models) can be efficiently used.

6 Highlights of the year

6.1 Awards

- Fabien Gandon won the best senior PC award at ESWC 2024.
- Best resource paper at ESWC2024 for the work on PyGraft [65] and best student research paper at ESWC2024 for a PhD student co-supervised by Pierre Monnin on the topic of semantically enhanced loss functions to learn graph embedding [64].
- Elena Cabrio was nominated for a lecture at College de France.
- Serena Villata received the “Victoires de la Recherche” award from the Ville de Nice.
- Outstanding Demo Award at the 27th European Conference on Artificial Intelligence (ECAI 2024) for the demo paper "PEACE: Providing Explanations and Analysis for Combating Hate Expressions" by Greta Damo, Nicolás Benjamín Ocampo, Elena Cabrio and Serena Villata [58]

7 New software, platforms, open data

7.1 New software

7.1.1 ACTA

Name: A Tool for Argumentative Clinical Trial Analysis

Keywords: Artificial intelligence, Natural language processing, Argument mining

Functional Description: Argumentative analysis of textual documents of various nature (e.g., persuasive essays, online discussion blogs, scientific articles) allows to detect the main argumentative components (i.e., premises and claims) present in the text and to predict whether these components are connected to each other by argumentative relations (e.g., support and attack), leading to the identification of (possibly complex) argumentative structures. Given the importance of argument-based decision making in medicine, ACTA is a tool for automating the argumentative analysis of clinical trials. The tool is designed to support doctors and clinicians in identifying the document(s) of interest about a certain disease, and in analyzing the main argumentative content and PICO elements. In 2024, ACTA has been integrated in a suite called ANTIDOTE, which collects the software results from the ANTIDOTE project (all partners).

URL: <http://antidote.i3s.unice.fr/acta/>

Contact: Serena Villata

7.1.2 ARViz

Name: Association Rules Visualization

Keyword: Information visualization

Scientific Description: ARViz supports the exploration of data from named entities knowledge graphs based on the joint use of association rule mining and visualization techniques. The former is a widely used data mining method to discover interesting correlations, frequent patterns, associations, or casual structures among transactions in a variety of contexts. An association rule is an implication of the form $X \rightarrow Y$, where X is an antecedent itemset and Y is a consequent itemset, indicating that transactions containing items in set X tend to contain items in set Y . Although the approach helps reduce and focus the exploration of large datasets, analysts are still confronted with the inspection of hundreds of rules in order to grasp valuable knowledge. Moreover, when extracting association rules from named entities (NE) knowledge graphs, the items are NEs that form antecedent \rightarrow consequent links, which the user should be able to cross to recover information. In this context, information visualization can help analysts visually identify interesting rules that are worthy of further investigation, while providing suitable visual representation to communicate the relationships between itemsets and association rules.

Functional Description: ARViz supports the exploration of thematic attributes describing association rules (e.g. confidence, interestingness, and symmetry) through a set of interactive, synchronized, and complementary visualisation techniques (i.e. a chord diagram, an association graph, and a scatter plot). Furthermore, the interface allows the user to recover the scientific publications related to rules of interest.

Release Contributions: Visualization of association rules within the scientific literature of COVID-19.

Publication: [hal-03292140](https://hal.archives-ouvertes.fr/hal-03292140)

Contact: Marco Alba Winckler

Participants: Aline Menin, Lucie Cadorel, Andrea Tettamanzi, Alain Giboin, Fabien Gandon, Marco Alba Winckler

7.1.3 Attune

Name: Attune - A Web-Based Digital Audio Workstation to Empower Cochlear Implant Users

Keywords: Web Application, Audio signal processing, Plug-in

Functional Description: Attune is an online software based on the Wam-Studio open-source digital audio workstation, adapted to help cochlear implant users perceive music more clearly. During multi-track listening, simple settings such as "clarity", "power", "attenuation" can be used. In reality, these settings control many parameters of sound processing plugins, which operate behind the scenes. The mapping between these plugins and the settings offered to users is carried out by researchers, using a dedicated graphical interface and a powerful macro management system.

Contact: Michel Buffa

Partner: CCRMA Lab, Stanford

7.1.4 CORESE-Core

Name: COnceptual REsource Search Engine - Core

Keywords: Semantic Web, RDF, RDFS, SPARQL, OWL, SHACL, Automated Reasoning, Validation, Interoperability, Linked Data, Knowledge Graphs, Knowledge Bases, Knowledge representation, Querying, Ontologies

Scientific Description: CORESE-Core is a library used in research to apply and evaluate Semantic Web standards and the algorithms they require. It is also the basis for proposing and prototyping extensions to these standards and their processing.

Functional Description: CORESE-Core is a library that implements and extends the Semantic Web standards established by the W3C, such as RDF, RDFS, SPARQL1.1 Query & Update, OWL RL, SHACL, and others.

This library offers a wide range of features for creating, manipulating, parsing, serializing, querying, reasoning and validating RDF data.

In addition, it offers advanced extensions such as STTL, SPARQL Rule and LDScript, which extend the functionality and processing capabilities of the data.

NB: CORESE-Core is a library derived from the earlier CORESE software.

Release Contributions: <https://github.com/Wimmics/corese/blob/master/CHANGELOG.md>

URL: <http://project.inria.fr/corese>

Contact: Remi Ceres

Participants: Remi Ceres, Fabien Gandon

7.1.5 CORESE-GUI

Name: COnceptual REsource Search Engine - Graphical User Interface

Keywords: GUI (Graphical User Interface), User Interfaces, Knowledge Bases, Knowledge Graphs, Knowledge graph, Knowledge representation, Ontologies, Linked Data, Validation, Automated Reasoning, SHACL, OWL, SPARQL, RDFS, RDF, Querying, Applications

Scientific Description: CORESE-GUI is a graphical user interface developed to interact with the CORESE-Core library. It provides users, especially those less experienced in programming, with an intuitive and visual access to the functionalities of CORESE-Core. This interface includes tools for visualizing semantic data, editing SPARQL queries, and monitoring data processing results. CORESE-GUI also serves as a platform for experimenting with new extensions and processing methods in the field of semantic web, thereby making these technologies more accessible to researchers and practitioners.

Functional Description: This desktop application allows the user to call up CORESE-Core features for creating, manipulating, parsing, serializing, querying, reasoning and validating RDF data.

The application enables direct use of Semantic Web languages standardized by the W3C, such as RDF and its syntaxes, RDFS, SPARQL1.1 Query & Update, OWL RL, SHACL, and others.

URL: <http://project.inria.fr/corese>

Contact: Remi Ceres

Participants: Remi Ceres, Fabien Gandon

7.1.6 CORESE-Server

Name: COnceptual REsource Search Engine - Server

Keywords: Server, Linked Data, Semantic Web, Ontologies, Knowledge Graphs, Knowledge Bases, RDF, RDFS, SPARQL, SHACL, Querying, Validation, Automated Reasoning

Scientific Description: This server version allows remote applications to access CORESE-Core functionalities for creating, manipulating, analyzing, serializing, querying, reasoning, and validating RDF data. The server facilitates remote use of W3C-standardized Semantic Web languages, such as RDF and its syntaxes, RDFS, SPARQL1.1 Query & Update, OWL RL, SHACL, and more.

Functional Description: This server version enables a remote application to call CORESE-Core's functions for creating, manipulating, analyzing, serializing, querying, reasoning and validating RDF data.

The server enables remote use of Semantic Web languages standardized by the W3C, such as RDF and its syntaxes, RDFS, SPARQL1.1 Query & Update, OWL RL, SHACL, and others.

URL: <https://project.inria.fr/corese/>

Contact: Remi Ceres

Participants: Remi Ceres, Fabien Gandon

7.1.7 CORESE-Command

Name: COnceptual REsource Search Engine - Command Line

Keywords: Command, RDF, RDFS, SPARQL, SHACL, Knowledge acquisition

Scientific Description: This command-line version of CORESE enables users to incorporate CORESE-Core functionalities into scripts, workflows, and consoles for creating, manipulating, analyzing, serializing, querying, reasoning, and validating RDF data. It allows direct use of W3C-standardized Semantic Web languages, such as RDF and its syntaxes, RDFS, SPARQL1.1 Query & Update, OWL RL, SHACL, and more.

Functional Description: This command-line version enables users to call CORESE-Core's functionality in scripts, workflows and console mode for the creation, manipulation, analysis, serialization, querying, reasoning and validation of RDF data.

The command enables direct use of W3C-standardized Semantic Web languages, such as RDF and its syntaxes, RDFS, SPARQL1.1 Query & Update, OWL RL, SHACL, and others.

URL: <https://project.inria.fr/corese/>

Contact: Remi Ceres

Participants: Remi Ceres, Fabien Gandon

7.1.8 CROBORA

Name: Crossing borders Archives. The circulation of images of Europe.

Keywords: Audiovisual, Data visualization

Functional Description: This platform gives access to 36 000 stock shots reused in the evening news of six national channels in France and Italy (France 2, Arte, TF1, Rai Uno, Rai Due and Canale 5) and the YouTube accounts of the European institutions between 2001 and 2021. The platform gives access four types of data: screenshots (one for each stock shot), metadata (for each stock shot), videos (news), and original metadata for each video. The platform integrates three visualization tools (Treemaps, Muvin, and ArViz) presenting patterns and relationships between records. The tool is available at <https://crobora.huma-num.fr/crobora?tab=0>

Contact: Marco Alba Winckler

7.1.9 Datalens

Keywords: Data visualization, Artificial intelligence

Functional Description: Datalens leverages custom network topologies, multi-faceted filters, and advanced visualization techniques to help users discover relevant datasets published online for their specific tasks. It harnesses the visualization capabilities of MGExplorer to enable a multi-perspective exploration of data. Currently, the tool supports navigation through datasets and models available on HuggingFace.

URL: <https://dataviz.i3s.unice.fr/datalens/>

Contact: Aline Menin

7.1.10 DBpedia

Name: DBpedia

Keywords: RDE, SPARQL

Functional Description: DBpedia is an international crowd-sourced community effort to extract structured information from Wikipedia and make this information available on the semantic Web as linked open data. The DBpedia triple stores then allow anyone to solve sophisticated queries against Wikipedia extracted data, and to link the different data sets on these data. The French chapter of DBpedia was created and deployed by Wimmics and is now an online running platform providing data to several projects such as: QAKIS, Izipedia, zone47, Sépage, HdA Lab., JocondeLab, etc.

Release Contributions: The new release is based on updated Wikipedia dumps and the inclusion of the DBpedia history extraction of the pages.

URL: <http://wiki.dbpedia.org/>

Contact: Fabien Gandon

Participants: Fabien Gandon, Elmahdi Korfed

7.1.11 Fuzzy labelling argumentation module

Name: Fuzzy labelling algorithm for abstract argumentation

Keywords: Artificial intelligence, Multi-agent, Knowledge representation, Algorithm

Functional Description: The goal of the algorithm is to compute the fuzzy acceptability degree of a set of arguments in an abstract argumentation framework. The acceptability degree is computed from the trustworthiness associated with the sources of the arguments.

Contact: Serena Villata

Participant: Serena Villata

7.1.12 GUsT-3D

Name: Guided User Tasks Unity plugin for 3D virtual reality environments

Keywords: 3D, Virtual reality, Interactive Scenarios, Ontologies, User study

Functional Description: We present the GUsT-3D framework for designing Guided User Tasks in embodied VR experiences, i.e., tasks that require the user to carry out a series of interactions guided by the constraints of the 3D scene. GUsT-3D is implemented as a set of tools that support a 4-step workflow to : (1) annotate entities in the scene with names, navigation, and interaction possibilities, (2) define user tasks with interactive and timing constraints, (3) manage scene changes, task progress, and user behavior logging in real-time, and (4) conduct post-scenario analysis through spatio-temporal queries on user logs, and visualizing scene entity relations through a scene graph.

The software also includes a set of tools for processing gaze tracking data, including: cleaning and synchronization of the data, calculation of fixations with I-VT, I-DT, IDTVR, IS5T, Remodnav, and IDVT algorithms, and visualization of the data (points of regard and fixations) in both real time and collectively.

URL: <https://project.inria.fr/creative3d/gust-3d/>

Publications: [hal-04102737](#), [hal-04446066](#), [hal-03635452](#)

Contact: Hui-Yin Wu

Participants: Hui-Yin Wu, Marco Alba Winckler, Lucile Sassatelli, Florent Robert

Partner: I3S

7.1.13 IndeGx

Keywords: Semantic Web, Indexation, Metadata

Functional Description: IndeGx is a framework for the creation of an index of a set of SPARQL endpoints. The framework relies only on available semantic web technologies and the index appears as an RDF database. The index is primarily composed of the self-description available in the endpoint. This original description is verified and expanded by the framework, using SPARQL queries.

Release Contributions: The previous version was a Java application coded with Apache Jena, this version uses an engine coded in Typescript with rdfliib, graphy, sparqljs, coupled with a Corese Server, in a docker application. - Treatment of endpoints in parallel -The automatic pagination of simple queries to avoid overwhelming SPARQL endpoints. - The usage of Corese as an interface with SPARQL endpoints to reduce missing data due to errors coming from incorrect standard compliance in distant SPARQL endpoints. - Rules are now expected to make heavy use of federated querying, with the SERVICE clause. - Possibility to define the application of several rules as a prerequisite to the application of another. - End of the difference between CONSTRUCT and UPDATE rules to differentiate between the application of local and distant queries. Only test queries are supposed to be SELECT, ASK, or CONSTRUCT. All action queries are expected to be UPDATE queries. - Possibility to define a set of rules as a pre-treatment or a post-treatment on the extracted data. In this case, the endpoint URL becomes the URL of the local corese server (not accessible from the outside of the docker) - Handling many different errors in the RDF format of data found in remote endpoints - Possibility of disabling the query logging of the framework - Possibility of using the query logging of the framework to avoid repeating rule application in case of an execution interruption - Integration of LDscript in rules possible.

We also offer two automatically refreshed catalogs, - The catalog of endpoints taken from numerous sources, updated daily - The catalog of endpoints and their statuses, refreshed hourly

URL: <https://github.com/Wimmics/IndeGx>

Publication: [hal-03946680v1](#)

Contact: Pierre Maillot

Participants: Fabien Gandon, Catherine Faron, Olivier Corby, Franck Michel

7.1.14 ISSA-pipeline

Name: Processing pipeline of the ISSA project

Keywords: Indexing, Semantic Web, Knowledge Graphs, NLP

Functional Description: See the description at <https://github.com/issa-project/issa-pipeline/tree/main/pipeline>

Release Contributions: Add bibliometric indicators and new visualizations. Details: <https://github.com/issa-project/issa-pipeline/compare/2.0...2.1>

URL: <https://github.com/issa-project/issa-pipeline>

Contact: Franck Michel

Participants: Anna Bobasheva, Franck Michel

7.1.15 ISSA Visualization Web Application

Keywords: Open Access, Data visualization, Knowledge graph, NLP

Functional Description: The ISSA project focuses on the semantic indexing of scientific publications in an open archive. The ISSA Visualization Web Application is a React and node.js based web application meant to search articles from the ISSA knowledge base using the rich semantics of the reference vocabularies, and provide a visualization of their metadata. This application consists of a frontend and a backend hosted on separate repositories: <https://github.com/issa-project/web-visualization/> <https://github.com/issa-project/web-backend>

URL: <https://github.com/issa-project/web-visualization/>

Contact: Franck Michel

Participant: Franck Michel

Partners: CIRAD, IMT Mines Alès

7.1.16 KartoGraphI

Functional Description: Website displaying a screenshot of the state of the Linked Data web according to the description retrieved by the IndeGx software

URL: <http://prod-dekalog.inria.fr/>

Publication: hal-03652865

Contact: Pierre Maillot

7.1.17 Licentia

Keywords: Right, License

Scientific Description: In order to ensure the high quality of the data published on the Web of Data, part of the self-description of the data should consist in the licensing terms which specify the admitted use and re-use of the data by third parties. This issue is relevant both for data publication as underlined in the “Linked Data Cookbook” where it is required to specify an appropriate license for the data, and for the open data publication as expressing the constraints on the reuse of the data would encourage the publication of more open data. The main problem is that data producers and publishers often do not have extensive knowledge about the existing licenses, and the legal terminology used to express the terms of data use and reuse. To address this open issue, we present Licentia, a suite of services to support data producers and publishers in data licensing by means of a user-friendly interface that masks to the user the complexity of the legal reasoning process. In particular, Licentia offers two services: i) the user selects among a pre-defined list those terms of use and reuse (i.e., permissions, prohibitions, and obligations) she would assign to the data and the system returns the set of licenses meeting (some of) the selected requirements together with the machine readable licenses’ specifications, and ii) the user selects a license and he/she can verify whether a certain action is allowed on the data released under such license. Licentia relies on the dataset of machine-readable licenses (RDF, Turtle syntax, ODRL vocabulary and Creative Commons vocabulary) available at <http://datahub.io/dataset/rdflicense>. We rely on the deontic logic to address the problem of verifying the compatibility of the licensing terms in order to find the license compatible with the constraints selected by the user. The need for licensing compatibility checking is high, as shown by other similar services (e.g., Licensius or Creative Commons Choose service). However, the advantage of Licentia with respect to these services is twofold: first, in these services compatibility is pre-calculated among a pre-defined and small set of licenses, while in Licentia compatibility is computed at runtime and we consider more than 50 heterogeneous licenses, second, Licentia provides a further service that is not considered by the others, i.e., it allows to select a license from our dataset and verify whether some selected actions are compatible with such license.

Functional Description: Licentia is a web service application with the aim to support users in licensing data. Our goal is to provide a full suite of services to help in the process of choosing the most suitable license depending on the data to be licensed.

The core technology used in our services is powered by the SPINdle Reasoner and the use of Defeasible Deontic Logic to reason over the licenses and conditions.

The dataset of RDF licenses we use in Licentia is the RDF licenses dataset where the Creative Commons Vocabulary and Open Digital Rights Language (ODRL) Ontology are used to express the licenses.

URL: <http://licentia.inria.fr/>

Contact: Serena Villata

Participant: Cristian Cardellino

7.1.18 Metadatamatic

Keywords: RDF, Semantic Web, Metadata

Functional Description: Website offering a form to generate in RDF the description of an RDF base.

URL: <https://wimmics.github.io/voidmatic/>

Contact: Pierre Maillot

Participants: Fabien Gandon, Franck Michel, Olivier Corby, Catherine Faron

7.1.19 MGExplorer

Name: Multivariate Graph Explorer

Keywords: Information visualization, Linked Data

Scientific Description: MGExplorer (Multidimensional Graph Explorer) allows users to explore different perspectives to a dataset by modifying the input graph topology, choosing visualization techniques, arranging the visualization space in meaningful ways to the ongoing analysis and retracing their analytical actions. The tool combines multiple visualization techniques and visual querying while representing provenance information as segments connecting views, which each supports selection operations that help define subsets of the current dataset to be explored by a different view. The adopted exploratory process is based on the concept of chained views to support the incremental exploration of large, multidimensional datasets. Our goal is to provide visual representation of provenance information to enable users to retrace their analytical actions and to discover alternative exploratory paths without losing information on previous analyses.

Functional Description: MGExplorer is an information visualization tool designed for exploring multivariate graphs, integrating various visualization techniques. It allows users to select and combine these techniques into a graph that traces the exploration path of a database. Developed with the D3.JS library, MGExplorer runs directly in a web browser. The tool is available online and can be customized using SPARQL queries created and managed within the LDViz software, which facilitates the creation, storage, and management of such queries. Additionally, MGExplorer can be integrated into any web project as an npm package, providing a modular solution for data visualization.

Release Contributions: MGExplorer is now available as a web component, making it easy to integrate into any web project via an npm package, accessible at <https://www.npmjs.com/package/mgexplorer>. It can be customized to visualize either local datasets or results from SPARQL queries.

URL: <https://dataviz.i3s.unice.fr/mgexplorer>

Publications: [hal-03292172](#), [hal-03404572](#), [hal-03404580](#)

Contact: Aline Menin

Participants: Aline Menin, Marco Alba Winckler, Olivier Corby

Partner: Universidade Federal do Rio Grande do Sul

7.1.20 Morph-xR2RML

Name: Morph-xR2RML

Keywords: RDF, Semantic Web, LOD - Linked open data, MongoDB, SPARQL

Functional Description: The xR2RML mapping language that enables the description of mappings from relational or non relational databases to RDF. It is an extension of R2RML and RML.

Morph-xR2RML is an implementation of the xR2RML mapping language, targeted to translate data from the MongoDB database, as well as relational databases (MySQL, PostgreSQL, MonetDB). Two running modes are available: (1) the graph materialization mode creates all possible RDF triples at once, (2) the query rewriting mode translates a SPARQL 1.0 query into a target database query and returns a SPARQL answer. It can run as a SPARQL endpoint or as a stand-alone application.

Morph-xR2RML was developed by the I3S laboratory as an extension of the Morph-RDB project which is an implementation of R2RML.

URL: <https://github.com/frmichel/morph-xr2rml/>

Publications: [hal-01207828](#), [hal-01330146](#), [hal-01280951](#)

Contact: Franck Michel

7.1.21 Muvin

Name: Multidimensional Visualization of Networks over Time

Keywords: Data visualization, LOD - Linked open data, Temporal Networks

Scientific Description: Muvin addresses the challenges of visualizing complex collaboration networks by implementing an incremental approach tailored for exploring co-authorship networks composed of multivariate entities distributed over time. Traditional representations of such networks can become visually cluttered, making it difficult to focus on relevant information. To tackle this, Muvin employs a focus+context technique, allowing users to zoom in on specific data points while maintaining an overview of the broader network. By enabling incremental data exploration and supporting multi-layered linked open data (LOD), Muvin effectively handles the complexity and scalability issues of collaboration networks. This approach intends to facilitate domain-specific tasks, such as identifying influential collaborators and understanding knowledge dissemination in co-authorship networks.

Functional Description: Muvin facilitates the exploration of a two-layer network that captures collaborations among entities such as researchers, artists, keywords, and more, as well as the temporal evolution of related elements, including scientific publications or songs. The tool adopts an incremental approach, enabling users to dynamically import data from a SPARQL endpoint into the exploration workflow. SPARQL queries can be created and adjusted on the fly using the LDViz query management tool, allowing users to experiment with different queries to address specific data-related questions. Developed with the D3.js library for visualization, Muvin is designed primarily for exploring data from knowledge graphs. The tool is accessible online at <https://dataviz.i3s.unice.fr/muvin>.

Release Contributions: This new version has been generalized to enable the exploration of collaboration networks among all types of entities, based on a wide variety of elements. It also allows the use of data from any SPARQL endpoint through custom queries, prepared directly within the LDViz tool (available at [<https://dataviz.i3s.unice.fr/ldviz>] (<https://dataviz.i3s.unice.fr/ldviz>)). This flexibility supports addressing specific questions across diverse and varied domains.

URL: <http://dataviz.i3s.unice.fr/muvin/>

Publication: [hal-03748134](https://hal.archives-ouvertes.fr/hal-03748134)

Contact: Aline Menin

Participants: Aline Menin, Marco Alba Winckler

7.1.22 Olivaw

Name: Ontology Long-lived Integration Via ACIMOV Workflow

Keywords: Ontologies, Ontology engineering, Semantic Web, Git svn, Linked Data, LOD - Linked open data, Web

Scientific Description: Olivaw proposes: (1) command lines that make an Acimov ontology development easier, (2) composite actions that can directly be called in workflows from any Acimov project, (3) a pre-commit hook that prevents mistakes from being pushed to an Acimov repository. The test reports are first represented using the EARL vocabulary and then exported in the markdown format to fit a github environment. A template repository also exists in order for an ontology project to begin with the accurate repository architecture, workflows and special files.

Functional Description: Agile and collaborative approaches to ontology development are crucial because they contribute to making them user-driven, up-to-date, and able to evolve alongside the systems they support, hence proper continuous validation tooling is required to ensure ontologies match these standards all along their development. We propose OLIVAW (Ontology Long-lived Integration Via ACIMOV Workflow), a tool supporting the ACIMOV methodology on GitHub. It relies on W3C Standards to assist the development of modular ontologies through GitHub Composite Actions, pre-commit hooks, or a command line interface. OLIVAW was tested on several ontology projects to ensure its usefulness, genericity and reusability. A template repository is available for a quick start. OLIVAW is published under the LGPL-2.1 license and archived on Software Heritage and Zenodo.

URL: <https://github.com/Wimmics/olivaw>

Publication: [hal-04823080](https://hal.archives-ouvertes.fr/hal-04823080)

Contact: Nicolas Robert

Partner: IMT - MINES Saint-Étienne

7.1.23 PEACE

Name: Providing Explanations and Analysis for Combating Hate Expressions

Keywords: Hate Speech Detection, Generating Explanations, Implicit Hate Speech, Subtle Hate Speech

Functional Description: PEACE is a web tool conceived to support content moderators in exploring and evaluating implicit and subtle hate speech on social media. It comprises three main functionalities: i) the exploratory analysis of hate speech messages characteristics (exploration), ii) the prediction of hatefulness (detection), and iii) the explanation of system predictions (explanation).

These functionalities incorporate not only a binary classification of whether a message is hateful (including explicit, implicit, and subtle messages), with a detailed explanation in natural language that clarifies why a message is considered hateful and an exploratory analysis of the message characteristics.

URL: <https://3ia-demos.inria.fr/peace/>

Contact: Elena Cabrio

7.1.24 plic2owl

Name: Plic2OWL: PlinianCore-to-OWL translation

Keywords: Biodiversity, Semantic Web, Knowledge acquisition, Ontologies

Functional Description: The Plinian Core vocabulary is a standard data model designed to share biological species level information. It is developed as an XML schema (XSD).

The Plinian Core ontology is a representation the XSD PlinianCore data model as an OWL ontology, to be used in RDF-based knowledge graphs.

This repository is a Python application that translates the Plinian Core XML schema into an OWL ontology. The output format is RDF Turtle.

URL: <https://github.com/Wimmics/plic2owl>

Contact: Franck Michel

Partners: GBIF.ES, Real Jardín Botánico, CSIC

7.1.25 RDFminer

Keywords: Evolutionary Algorithms, Semantic Web, Web API, Dashboard

Functional Description: RDFminer is an open source Web application to automatically discover SHACL shapes through an evolutionary process. It takes an RDF data graph as input, from which shapes are mined and assessed using a probabilistic validation framework. The user can interact with RDFminer through a dashboard where they can launch and monitor the mining of shapes, and analyse the results in real time. Github: <https://github.com/Wimmics/RDFminer/>

URL: <http://ns.inria.fr/rdfminer>

Publication: [hal-04566981v1](https://hal.archives-ouvertes.fr/hal-04566981v1)

Contact: Remi Felin

Participants: Remi Felin, Thu Nguyen, Andrea Tettamanzi, Catherine Faron, Fabien Gandon

7.1.26 SciLEX

Name: Science Literature Exploration

Keywords: Textmining, Systematic review, Collaborative science, Linked Data

Functional Description: Scilex is a tool allows to start a scientific paper collect to analyse the state of art of a given domain. It also allows the annotation, the enrichment as well as the analysis of the results.

Publication: [hal-04175511](https://hal.archives-ouvertes.fr/hal-04175511)

Contact: Celian Ringwald

Participants: Celian Ringwald, Anaïs Ollagnier

7.1.27 SPARQL Micro-services

Name: SPARQL micro-services

Keywords: Web API, RDF, SPARQL, Semantic Web, Knowledge Graphs

Functional Description: The approach leverages the micro-service architectural principles to define the SPARQL Micro-Service architecture, aimed at querying Web APIs using SPARQL. A SPARQL micro-service is a lightweight SPARQL endpoint that typically provides access to a small, resource-centric graph. Furthermore, this architecture can be used to dynamically assign dereferenceable URIs to Web API resources that do not have URIs beforehand, thus literally “bringing” Web APIs into the Web of Data. The implementation supports a large scope of JSON-based Web APIs, may they be RESTful or not.

Release Contributions: Various improvements, updates and bug fixes.

URL: <https://github.com/frmichel/sparql-micro-service>

Publications: [hal-02060966](#), [hal-01722792](#), [hal-01947589](#), [hal-02168164](#)

Contact: Franck Michel

7.1.28 wam-studio

Keywords: Web Application, Web API, Audio signal processing

Functional Description: WAM Studio is an open source online Digital Audio Workstation (DAW) that takes advantage of a number of APIs and standard W3C technologies, such as Web Audio, WebAssembly, Web Components, Web Midi, Media Devices and more. WAM Studio is also based on the Web Audio Modules (WAM) standard, which was designed to facilitate the development of interoperable audio plug-ins (effects, virtual instruments, virtual piano keyboards as controllers, etc.) - a kind of "VSTs for the Web". DAWs are feature-rich software programs, and therefore particularly complex to develop in terms of design, implementation, performance and ergonomics. Today, the majority of online DAWs are commercial, while the only open source examples lack functionality (no plug-in support, for example) and do not take advantage of the recent possibilities of web browsers (such as WebAssembly). WAM Studio was designed as a technology demonstrator to promote the possibilities offered by recent innovations proposed by the W3C. Developing it was a challenge, as we had to take into account the limitations of sandboxed and constrained environments such as Web browsers, and compensate for latency when we cannot know what hardware is being used, etc.). An online demo and a GitHub repository for the source code are available (<https://wam-studio.i3s.univ-cotedazur.fr/>).

Contact: Michel Buffa

7.1.29 WebAudio tube guitar amp sims CLEAN, DISTO and METAL MACHINES

Name: Tube guitar amplifier simulators for Web Browser : CLEAN MACHINE, DISTO MACHINE and METAL MACHINE

Scientific Description: This software is one of the only ones of its kind to work in a web browser. It uses "white box" simulation techniques combined with perceptual approximation methods to provide a quality of guitar playing in hand comparable to the best existing software in the native world.

Functional Description: Software programs for creating real-time simulations of tube guitar amplifiers that behave most faithfully like real hardware amplifiers, and run in a web browser. In addition, the generated simulations can run within web-based digital audio workstations as plug-ins. The "CLEAN MACHINE" version specializes in the simulation of acoustic guitars when playing electric guitars. The DISTO machine specializes in classic rock tube amp simulations, and METAL MACHINE targets metal amp simulations. These programs are one of the results of the ANR WASABI project.

Release Contributions: First stable version, delivered and integrated into the ampedstudio.com software. Two versions have been delivered: a limited free version and a commercial one.

Publications: [hal-01721463](#), [hal-01893681](#), [hal-02337828](#), [hal-03087768](#), [hal-01721483](#), [hal-01735478](#), [hal-02366725](#), [hal-02557901](#), [hal-01589330](#), [hal-03087763](#), [hal-01893660](#), [hal-01589229](#)

Contact: Michel Buffa

Participant: Michel Buffa

7.1.30 WheatGenomicsSLKG Visualization Web Application

Keywords: Open Access, Data visualization, Knowledge graph, NLP

Functional Description: The Wheat Genomics Scientific Literature Knowledge Graph (WheatGenomicsSLKG) is a FAIR knowledge graph that exploits the Semantic Web technologies to integrate information about Named Entities (NE) extracted automatically from a corpus of PubMed scientific articles on wheat genetics and genomics.

The WheatGenomicsSLKG Visualization Web Application is a React and node.js based web application meant to search articles from the WheatGenomicsSLKG using the rich semantics of the reference vocabularies, and provide a visualization of their metadata. This application consists of a frontend and a backend hosted on separate repositories: <https://github.com/Wimmics/wheatgenomicsslkg-web-visualization/> <https://github.com/Wimmics/wheatgenomicsslkg-web-backend/>

Contact: Franck Michel

7.1.31 Zoomathia KG Pipeline

Name: Automatic annotation of an ancient zoological corpus

Keywords: Zoology, NLP, Semantic annotation, Semantic Web

Functional Description: This project provides a text processing pipeline for a corpus of texts on animals compiled within the framework of the Zoomathia GDRI. The web interface allows researchers to explore the corpus via a search for works by concept, explore a selected work while visualizing the concepts annotating each of its parts, and visualize the results of queries implementing competency questions on a selected work from the corpus.

URL: <https://univ-cotedazur.fr/recherche-innovation/structures-de-recherche/academies-dexcellence/academie-dexcellence-homme-idees-et-milieus/projets-de-recherche/projets-2020-2024/automazoo-annotation-automatique-dun-corpus-zoologique-ancien>

Contact: Catherine Faron

Partners: Université de Nice Sophia Antipolis (UNS), CEPAM (Cultures, Environnements, Préhistoire, Antiquité, Moyen Âge)

7.2 Open data

CyberAgressionAdo

Contributors: Anaïs Ollagnier, Elena Cabrio, Serena Villata

Description: The *CyberAgressionAdo-v1* dataset comprises instances of aggressive multiparty chats in French, gathered through a role-playing game conducted in high schools. This dataset is built upon scenarios that emulate cyber aggression situations prevalent among teenagers, addressing sensitive topics like ethnic origin, religion, or skin color. The recorded conversations have undergone meticulous annotation, taking into account various facets, including participant roles, the occurrence of hate speech, the nature of verbal abuse within the messages, and the identification of humor devices like sarcasm or irony in utterances.

Dataset PID (DOI,...): 10.5281/zenodo.14770265

Project link: [CyberAgressionAdo-v1](#), [CyberAgressionAdo-V2](#), [CyberAgressionAdo-large](#)

Publications: [70]

Contact: Anaïs Ollagnier

Release contributions: *CyberAgressionAdo-V2* uses a multi-label, fine-grained tagset marking the discursive role of exchanged messages as well as the context in which they occur – for instance, attack (ATK), defend (DFN), counterspeech (CNS), abet/instigate (AIN), gaslight (GSL), etc. The *CyberAgressionAdo-Large* dataset builds on our previously published work on *CyberAgressionAdo*, an open-access French dataset created to support research on online hate detection within multiparty conversations. In this extended version, the dataset size has nearly doubled, growing from 19 to 36 aggressive multiparty conversations collected through role-playing simulations in schools. This expansion enables a more in-depth analysis across a broader spectrum of cyber aggression themes.

ElecDeb60to20

Contributors: Serena Villata, Elena Cabrio, Pierpaolo Goffredo

Description: The ElecDeb60to20 dataset is built from the official transcripts of the televised presidential debates in the US from 1960 until 2020, from the website of the Commission on Presidential Debates (CPD). These political debates are manually annotated with argumentative components (claim, premise) and relations (support, attack). In addition, it also includes the annotation of fallacious arguments, based on the following 6 classes of fallacies: ad hominem, appeal to authority, appeal to emotion, false cause, slogan, slippery slope.

Dataset PID (DOI,...): 10.18653/v1/2023.emnlp-main.684

Project link: [ElecDeb60to20](#)

Publications: [117]

Contact: Serena Villata

Release contributions: (first release)

TAXREF-LD: Knowledge Graph of the French taxonomic registry

Contributors: Franck Michel, Catherine Faron

Description: TAXREF-LD is a Linked Data knowledge graph representing TAXREF, the French national taxonomical register for fauna, flora and fungus, that covers mainland France and overseas territories. TAXREF-LD is a joint initiative of the UMS Patrinat of the National Museum of Natural History, and the I3S laboratory, University Côte d’Azur, Inria, CNRS.

Dataset PID (DOI,...): DOI:10.5281/zenodo.12733630

Project link: [TAXREF-LD](#), [10.5281/zenodo.12733630](#)

Publications: [119]

Contact: Franck Michel

Release contributions: version 17.0 implements new SKOS collections and better management of vernacular. See full description at <https://github.com/frmichel/taxref-ld/blob/master/CHANGELOG.md>

WheatGenomicsSLKG

Contributors: Nadia Yacoubi Ayadi, Franck Michel, Catherine Faron

Description: Wheat Genomics Scientific Literature Knowledge Graph is a FAIR knowledge graph that exploits the Semantic Web technologies to integrate information about Named Entities (NE) extracted automatically from a corpus of PubMed scientific papers on wheat genetics and genomics. This work is supported by the French National Research Agency under grant ANR-18-CE23-0017 (project [D2KAB](#)).

Dataset PID (DOI,...): DOI:10.5281/zenodo.10420888

Project link: [WheatGenomicsSLKG, 10.5281/zenodo.10420888](#)

Publications: [\[41\]](#)

Contact: Franck Michel

Release contributions: (first release)

Pharmacogenomics datasets for Ontology Matching

Contributors: Pierre Monnin

Description: These datasets constitute benchmarks to evaluate Ontology Matching algorithms on a complex structure-based instance matching task from the domain of pharmacogenomics. Pharmacogenomics involves n -ary tuples representing so-called “pharmacogenomic relationships” and their components of three distinct types: drugs, genetic factors, and phenotypes. The goal resides in matching such tuples. These datasets were extracted from the [PGxLOD](#) knowledge graph.

Dataset PID (DOI,...): DOI:10.5281/zenodo.8419361

Project link: [PGxLOD, 10.5281/zenodo.8419361](#)

Contact: Pierre Monnin

Release contributions: this is the first published version.

DBpedia.fr : French chapter of the DBpedia knowledge graph dataset

Contributors: Fabien Gandon, Franck Michel, Célian Ringwald,

Description: The [DBpedia.fr](#) project ensures the creation and maintenance of a French chapter of the DBpedia knowledge base a crowd-sourced community effort to extract structured content from the information created in various Wikimedia projects. Statistics indicate very high usage rate: the server processed 1.8+ billion queries over the year. This represents a 3.86 million daily average and 32.5 million daily max.

Dataset PID (DOI,...):

Project link: <https://fr.dbpedia.org/>

Contact: Célian Ringwald

Release contributions: No new release was done this year but we carried out continuous monitoring and support to ensure a high-availability service.

FALCON: Fallacies in COVID-19 Network-based dataset

Contributors: Mariana Chaves, Elena Cabrio, Serena Villata.

Description: The FALCON dataset is a collection of tweets related to the COVID-19 pandemic and politically associated discussions annotated with 6 fallacy categories: *loaded language*, *appeal to fear*, *appeal to ridicule*, *hasty generalization*, *ad hominem*, and *false dilemma*. Annotations are provided at the tweet level and in a multi-label format, meaning that a tweet can be associated with more than one fallacy category. The dataset includes an underlining graph structure that can be used to model the relationships between the fallacies.

Dataset PID (DOI,...): Not available yet.

Project link: [falcon-fallacy-classification](#)

Publications: [57]

Contact: Mariana Chaves

Release contributions: (first release)

8 New results

8.1 User Modeling and Designing Interaction

8.1.1 Interaction with extended reality

Participants: Aline Menin, Clément Quéré, Florent Robert, Marco Winckler.

Affordances are crucial tools for supporting interaction in Mixed, Virtual, and Augmented Reality (XR) environments. They define how the system communicates to the user the actions needed to perform a task. Given the lack of consensus and standards in the literature on describing affordances in XR, we explored how affordances are used as a strategy to encourage discovery and engagement within XR environments. We gathered and analyzed studies that explain affordance concepts and conduct experiments in 3D environments, highlighting the affordances applied. Our analysis mapped the use of affordances across eight dimensions: environment type, modality, location, temporal aspect, learning aspect, task category, domain tasks, and affordance type [80].

In particular, Mixed Reality (MR) offers a great potential for annotating physical spaces, allowing users to interact with the real world through digital annotations. Annotations serve various functions, such as providing descriptions, offering assessments, or both, and play a key role in daily life for summarizing and emphasizing content. MR facilitates these annotations by integrating them with the physical environment through head-mounted displays (HMDs), making interactions more seamless.

In fields like real estate, healthcare, and entertainment, MR-driven annotations enhance user experiences, offering immersive, informative, and collaborative benefits. In this context, we proposed and developed an annotation system called HandyNotes, designed to enrich physical objects with semantic information using MR headsets and multiple input methods (audio, gestures, text) [72]. To support seamless interaction with the real-world, we proposed an extended hand menu system for intuitive access to tools, an asynchronous note-taking process for sharing annotations. The lessons learned during the development of HandyNotes contribute to more natural and meaningful MR interactions. The tool is currently being enhanced to facilitate data visualization and exploration through annotations in Virtual Reality (VR). Additionally, we aim to simplify the storage, access, and sharing of produced annotations using semantic web technologies, allowing for easy identification of real-world entities and linking them to existing data on the web.

8.1.2 Visual exploration of datasets and models for AI

Participants: Aline Menin, Anaïs Ollagnier.

The rapid expansion of publicly available textual resources, such as lexicons, thesauri, and domain-specific corpora, presents challenges for data practitioners in efficiently identifying the most relevant resources for their tasks. While repositories are emerging to catalog and share these resources, they often lack the advanced search and exploration capabilities users need. Most search methods employ keyword-based queries and/or metadata filtering techniques, which typically return a raw list of possible interesting resources that match the search. This approach is mainly limited by (i) requiring prior knowledge of existing resources to craft effective queries and (ii) failing to reveal potential connections between resources. In response, we present *DataLens*, a comprehensive web-based platform that integrates faceted search with advanced information visualization techniques to improve resource discovery. Our solution features a filtering-based search, a network-based visualization, and a chained views approach to facilitate data exploration from diverse perspectives. *DataLens* currently manages a collection of 212,753 resources spanning eight distinct research modalities in machine learning. The tool is available online at [DataLens](#).

8.1.3 Interactive WebAudio applications

Participants: Michel Buffa, Shihong Ren.

During the WASABI ANR research project (2017-2020), we built a 2M song database made of metadata collected from the Web of Data and from the analysis of song lyrics of the audio files provided by Deezer. This dataset is still exploited by current projects inside the team (in particular by the PhD of Maroua Tickat). Other initiatives closely related to the WASABI datasets include several Web Audio interactive applications and frameworks. The Web Audio Modules 2.0, a WebAudio plugin standard for developing high performance plugins in the browser published for the first time in 2021, has encountered a large adoption by researchers and developers. The open source Wam-Studio Digital Audio Workstation developed by Michel Buffa and Antoine Vidal-Mazuy led to a scientific collaboration with the MERI team from the CCRMA Laboratory in Stanford. The team is also participating to the ANR DOTS project, that aims to produce distributed music performances using Web Audio based infrastructure.

We also developed new methods for real-time tube guitar amplifier simulations that run in the browser. Some of these results are still unique in the world as in 2023, and have been acclaimed by several awards in international conferences. The guitar amp simulations are now commercialized by the CNRS SATT service and are available in the online collaborative Digital Audio Workstation *ampedstudio*. Some other tools we designed are linked to the WASABI knowledge base, that allow, for example, songs to be played along with sounds similar to those used by artists. An ongoing PhD proposes a visual language for music composers to create instruments and effects linked to the WASABI corpus content and a research collaboration with the Shanghai conservatory of Music is also being pursued about the generation of music on the Web with real-time Brain Wave.

More recently, we started some research work on the Musical Metaverse, a concept that refers to an immersive virtual space dedicated to musical activities. It is an extension of the "Metaverse" concept, which defines a set of interconnected virtual worlds where users can interact with each other and with digital objects. We focused on the design of a persistent, real-time multi-participant immersive world for shared music installation creation, exploiting recent W3C web standards such as Web Audio, Web MIDI, WebXR, WebGL, WebGPU, WebAssembly, WebRTC now implemented in the web browsers of the most common VR/XR headsets available on the market. [73, 51, 52, 53, 54, 101, 102].

8.1.4 KartoGraphI: Drawing a Map of Linked Data

Participants: Pierre Maillot, Olivier Corby, Catherine Faron, Fabien Gandon, Franck Michel.

A large number of semantic Web knowledge bases have been developed and published on the Web. To help the user identify the knowledge bases relevant for a given problem, and estimate their usability, we propose a declarative indexing framework and an associated visualization Web application, KartoGraphI. It provides an overview of important characteristics for more than 400 knowledge bases including, for instance, dataset location, SPARQL compatibility level, shared vocabularies, etc. [118]

8.2 Wisnote: a method and tool support for annotation of Web resources

Participants: Maroua Tikat, Aline Menin, Marco Winckler, Michel Buffa.

Within an era characterized by the rapid dissemination of digital information, annotations surface as a potent instrument for infusing content with profound depth and precision. Annotation might assume diverse forms (ex. text notes, audio records, drawings, marginalia, etc.) and connect to various target document support (including Web documents, multimedia documents, graphs, etc.). In order to investigate the use of annotations in visual analytical processes we have devised and implemented the Wisnote approach which is delivered with an eponym tool support. The primary goal of Wisnote is to empower users with an annotation tool allowing them to create personal (and/or collaborative) records of their experience with Web contents. Wisnote was deployed in two versions: the first one allows to annotate DOM elements (ex. text, images, HTML tags, etc) in Web pages; this version is mainly aimed at supporting knowledge extraction and sense making of Web contents while users browse the Web. The second version allows to annotate a visual dashboard that shows a knowledge graph created from contents extracted from the Web; this version is aimed at exploring the use of annotations. In order to investigate the use of the two versions of the tool, the study adopted a mixed-methods approach, combining quantitative metrics with qualitative user feedback, to evaluate the usability and to understand the uses for annotations.

8.3 Communities and Social Interactions Analysis

8.3.1 Online Hate Detection in Conversational Data

Participants: Elena Cabrio, Serena Villata, Anais Ollagnier.

Online harassment is prevalent on social networks, and the increasing frequency of anti-social behavior has compelled platform hosts to seek innovative solutions. Most research efforts aimed at addressing online harassment focus on social networks such as Twitter and Instagram. However, recent studies highlight messaging platforms and chat rooms as primary venues for cyberbullying, especially among teenagers. Due to strict data collection policies imposed by major social media platforms, there has been limited research into this phenomenon in multi-party settings. The recent availability of resources simulating online aggression among teens on private messaging platforms has begun to address this gap.

A conference paper presenting a comprehensive study of representation learning methods to address this phenomenon in these contexts is currently under submission. The paper explores various representation learning strategies (including both text- and graph-based methods) and their combinations to tackle multiple hate-related sub-tasks within conversations. This paper makes three key contributions: (1) benchmarking computational models for multi-party dialogue focused on addressing online hate, (2) expanding evaluation considerations for hate-related sub-tasks, and (3) investigating online hate at the linguistic level of pragmatics.

In parallel, we have expanded upon our previously published work on *CyberAgressionAdo*, an open-access French dataset designed to support research on online hate detection within multiparty conversations. In this extended version, the dataset size has nearly doubled, increasing from 19 to 36 aggressive multiparty conversations collected through role-playing simulations in schools. This expansion facilitates a more comprehensive analysis across a broader range of cyber aggression themes. [CyberAgression-Large github URL](#).

A journal paper consolidating all the methodological protocols for the collection and annotation of these data is currently under publication. The contribution of this paper lies in the significant improvements made to the annotation process, including enhanced guidelines and a two-phase inter-annotator agreement experiment to ensure greater consistency and clarity in labeling. Furthermore, the paper introduces an adaptation of the Weirdness Index to analyze lexical-level disagreements among annotators, offering a novel perspective on understanding divergent interpretations in annotation tasks.

8.3.2 Unveiling the Hate: Generating Faithful and Plausible Explanations for Implicit and Subtle Hate Speech Detection

Participants: Greta Damo, Nicolas Ocampo, Elena Cabrio, Serena Villata.

In today's digital age, the huge amount of abusive content and hate speech on social media platforms presents a significant challenge. Natural Language Processing (NLP) methods have focused on detecting explicit forms of hate speech, often overlooking more nuanced and implicit instances. To address this gap, our paper aims to enhance the detection and understanding of implicit and subtle hate speech. More precisely, we propose a comprehensive approach combining prompt construction, free-text generation, few-shot learning, and fine-tuning to generate explanations for hate speech classification, with the goal of providing more context for content moderators to unveil the actual nature of a message on social media [59].

8.3.3 PEACE: Providing Explanations and Analysis for Combating Hate Expressions

Participants: Greta Damo, Nicolas Ocampo, Elena Cabrio, Serena Villata.

The increasing presence of hate speech (HS) on social media poses significant societal challenges. While efforts in the Natural Language Processing community have focused on automating the detection of explicit forms of HS, subtler and indirect expressions often go unnoticed. This demo presents PEACE, a novel tool that, besides detecting if a social media message contains explicit or implicit HS, also generates detailed natural language explanations for such predictions. More specifically, PEACE addresses three main challenging tasks: i) exploring the characteristics of HS messages, ii) predicting hatefulness, and iii) elucidating the reasoning behind system predictions. A REST API is also provided to exploit the tool's functionalities [58].

8.3.4 Is Safer Better? The Impact of Guardrails on the Argumentative Strength of LLMs in Hate Speech Countering

Participants: Helena Bonaldi, Greta Damo, Nicolas Ocampo, Elena Cabrio, Serena Villata.

The potential effectiveness of counterspeech as a hate speech mitigation strategy is attracting increasing interest in the NLG research community, particularly towards the task of automatically producing it. However, automatically generated responses often lack the argumentative richness which characterizes expert-produced counterspeech. In this work, we focus on two aspects of counterspeech generation to

produce more cogent responses. First, by investigating the tension between helpfulness and harmlessness of LLMs, we test whether the presence of safety guardrails hinders the quality of the generations. Secondly, we assess whether attacking a specific component of the hate speech results in a more effective argumentative strategy to fight online hate. By conducting an extensive human and automatic evaluation, we show how the presence of safety guardrails can be detrimental also to a task that inherently aims at fostering positive social interactions. Moreover, our results show that attacking a specific component of the hate speech, and in particular its implicit negative stereotype and its hateful parts, leads to higher-quality generations [50].

8.3.5 Argumentation Quality Assessment

Participants: Elena Cabrio, Serena Villata.

We investigated, in [81], the main approaches presented in the literature for the automatic assessment of natural language argumentation. The computational treatment of arguments on controversial issues has been subject to extensive NLP research, due to its envisioned impact on opinion formation, decision making, writing education, and the like. A critical task in any such application is the assessment of an argument's quality - but it is also particularly challenging. We start from a brief survey of argument quality research, where we identify the diversity of quality notions and the subjectiveness of their perception as the main hurdles towards substantial progress on argument quality assessment. We argue that the capabilities of instruction-following large language models (LLMs) to leverage knowledge across contexts enable a much more reliable assessment. Rather than just fine-tuning LLMs towards leaderboard chasing on assessment tasks, they need to be instructed systematically with argumentation theories and scenarios as well as with ways to solve argument-related problems. We discuss the real-world opportunities and ethical issues emerging thereby.

8.3.6 DISPUTool 2.0: A Modular Architecture for Multi-Layer Argumentative Analysis of Political Debates

Participants: Pierpaolo Goffredo, Elena Cabrio, Serena Villata.

Political debates are one of the most salient moments of an election campaign, where candidates are challenged to discuss the main contemporary and historical issues in a country. These debates represent a natural ground for argumentative analysis, which has always been employed to investigate political discourse structure and strategy in philosophy and linguistics. In this paper, we present DISPUTool 2.0, an automated tool which relies on Argument Mining methods to analyze the political debates from the US presidential campaigns to extract argument components (i.e., premise and claim) and relations (i.e., support and attack), and highlight fallacious arguments. DISPUTool 2.0 allows also for the automatic analysis of a piece of a debate proposed by the user to identify and classify the arguments contained in the text. A REST API is provided to exploit the tool's functionalities [17].

The new version of the DISPUTool demo is publicly accessible at 3ia-demos.inria.fr/disputool/.

8.3.7 Argument-based Detection and Classification of Fallacies in Political Debates

Participants: Pierpaolo Goffredo, Mariana Eugenia Chaves Espinoza, Elena Cabrio, Serena Villata.

Fallacies are arguments that employ faulty reasoning. Given their persuasive and seemingly valid nature, fallacious arguments are often used in political debates. Employing these misleading arguments in politics can have detrimental consequences for society, since they can lead to inaccurate conclusions

and invalid inferences from the public opinion and the policymakers. Automatically detecting and classifying fallacious arguments represents therefore a crucial challenge to limit the spread of misleading or manipulative claims and promote a more informed and healthier political discourse. Our contribution to address this challenging task is twofold. First, we extend the ElecDeb60To16 dataset of U.S. presidential debates annotated with fallacious arguments, by incorporating the most recent Trump-Biden presidential debate. We include updated tokenlevel annotations, incorporating argumentative components (i.e., claims and premises), the relations between these components (i.e., support and attack), and six categories of fallacious arguments (i.e., Ad Hominem, Appeal to Authority, Appeal to Emotion, False Cause, Slippery Slope, and Slogans). Second, we perform the twofold task of fallacious argument detection and classification by defining neural network architectures based on Transformers models, combining text, argumentative features, and engineered features. Our results show the advantages of complementing transformer-generated text representations with non-textual features [18].

8.3.8 FALCON: A multi-label graph-based dataset for fallacy classification in the COVID-19 infodemic

Participants: Mariana Eugenia Chaves Espinoza, Elena Cabrio, Serena Villata.

Fallacies are arguments that seem valid but contain logical flaws. During the COVID-19 pandemic, they played a role in spreading misinformation, causing confusion and eroding public trust in health measures. Therefore, there is a critical need for automated tools to identify fallacies in media, which can help mitigate harmful narratives in future health crises. We present two key contributions to address this task. First, we introduce FALCON, a multi-label, graph-based dataset containing COVID-19-related tweets. This dataset includes expert annotations for six fallacy types—*loaded language*, *appeal to fear*, *appeal to ridicule*, *hasty generalization*, *ad hominem*, and *false dilemma*—and allows for the detection of multiple fallacies in a single tweet. The dataset’s graph structure enables analysis of the relationships between fallacies and their progression in conversations. Second, we evaluate the performance of language models on this dataset and propose a dual-transformer architecture that integrates engineered features. Beyond model ranking, we conduct statistical analyses to assess the impact of individual features on model performance [57].

8.3.9 Augmenting participation, co-creation, trust and transparency in Deliberative Democracy at all scales through Argument Mining

Participants: Cristian Cardellino, Sofiane Elguendouze, Elena Cabrio, Serena Villata.

In the context of the ORBIS project ¹, Argument Mining (AM) holds significant potential to support AI-enhanced democratic deliberation and decision-making. The integration of AM methods into the BCause platform, a “Structured and Decentralised Discussion System for Distributed Decision Making” ² implied a specific approach to AM, since discourses are structured. BCause data presents a distinct format, where discussions are organized at the statement level. Each statement (i.e., an argumentative sentence articulated by a participant) is either a position (a stance taken on a particular argumentative topic) or an argument that either supports or attacks a given position. This data limitation and the contextual features of BCause prompted a shift towards refining AM models for both argument component detection and relation classification. The classical AM pipeline has been adjusted to focus on what we now refer to as ‘statement-level’ AM. Similar to traditional AM, the process is carried out in two key steps: (1) Statement Classification, where each statement is categorized as either a position, a supporting argument, or an attacking argument, and (2) Statement Relation Classification, where the specific positions targeted by specific supporting or attacking arguments are identified (linked to each other). Further enhancement and refinement of relation classification is achieved through iterative testing within pilot settings, and

¹ORBIS Web Site

²BCause Web Site

feedback from both pilot organizations and participants to their deliberative experiments as end-users. In parallel, a more classical AM approach has been experimented [103], revising the model to handle the formats of deliberative discourse to process the pilot data effectively.

8.3.10 Controversy and influence in the Ukraine war: a study of argumentation and counter-argumentation through Artificial Intelligence

Participants: Mariana Eugenia Chaves Espinoza, Sofiane Elguendouze, Elena Cabrio, Serena Villata.

To understand the impact of controversies on armed conflicts, in particular the one between Russia and Ukraine, the project CIGAIA³ is focusing on designing and implementing an Artificial Intelligence algorithm for the automatic analysis of argumentation in these controversies in both English and French newspaper articles. This work is based on the joint creation of an annotated dataset from an initial mapping, allowing automatic argumentation analysis to identify and classify controversies. At its current status, the work that has been achieved covers the data crawling and cleaning from the news articles and the preparation of the annotation guidelines to be used for annotating the data with argumentative components on prolific platform⁴.

8.3.11 Argument-structured Justification Generation for Explainable Fact-checking

Participants: Xiaou Wang, Elena Cabrio, Serena Villata.

Justification production is a central task in automated fact-checking, and most studies cast this task as summarization. However, the majority of previous studies presume the availability of human-written fact-checking articles, which is unrealistic in practice. In this work, we address this issue by proposing a novel approach to generate argument-based justifications to improve fact-checking. Our contribution is threefold. First, our extensive experimental setting shows that, despite lower ROUGE scores, our argument-structured summarizer produces summaries leading to better claim verification performance than the state-of-the-art summarizer in fact-checking on three different benchmarks for this task. Second, our jointly-trained summarization and evidence retrieval system outperforms the state-of-the-art method on ExClaim, the only dataset where no human-written fact-checking articles are provided during verification of news claims. Third, we show that integrating attackability evaluation into the training process of the summarizer significantly reduces hallucinated argument relations, leading to more reliable and trustworthy justification generation [82].

8.3.12 Study of the Attention Economy, its detrimental impacts and leads for regulation

Participants: Franck Michel, Fabien Gandon.

During the last two decades, leveraging research in psychology, sociology, neuroscience and other domains, Web platforms have brought the process of capturing attention to an unprecedented scale. With the initial commonplace goal of making targeted advertising more effective, the generalization of attention-capturing techniques and their use of cognitive biases and emotions have multiple detrimental side effects such as polarizing opinions, spreading false information and threatening public health, economies and democracies.

Aware of the problems raised and our responsibility as a community, since 2023 we have initiated a work meant to warn the computer science community and call for regulation. We brought together

³[link to CIGAIA project](#)

⁴[Prolific annotation platform](#)

contributions from a wide range of disciplines (psychology, sociology, neuroscience, politics, legal domain, computer science, education etc.) to analyze current practices and consequences thereof. We published this work in 2024 at the AI, Ethics and Society conference (AIES) [68] in an article that provides a set of propositions and principles that could be used to drive further works, and to call for actions against these practices competing to capture our attention on the Web.

8.3.13 CasiMedicos-Arg: A Medical Question Answering Dataset Annotated with Explanatory Argumentative Structures

Participants: Ekaterina Sviridova, Elena Cabrio, Serena Villata.

Explaining Artificial Intelligence (AI) decisions is a major challenge nowadays in AI, in particular when applied to sensitive scenarios like medicine and law. However, the need to explain the rationale behind decisions is a main issue also for human-based deliberation as it is important to justify why a certain decision has been taken. Resident medical doctors for instance are required not only to provide a (possibly correct) diagnosis, but also to explain how they reached a certain conclusion. Developing new tools to aid residents to train their explanation skills is therefore a central objective of AI in education. In this paper, we follow this direction, and we present, to the best of our knowledge, the first multilingual dataset for Medical Question Answering where correct and incorrect diagnoses for a clinical case are enriched with a natural language explanation written by doctors. These explanations have been manually annotated with argument components (i.e., premise, claim) and argument relations (i.e., attack, support), resulting in the Multilingual CasiMedicos-Arg dataset which consists of 558 clinical cases in four languages (English, Spanish, French, Italian) with explanations, where we annotated 5021 claims, 2313 premises, 2431 support relations, and 1106 attack relations. We conclude by showing how competitive baselines perform over this challenging dataset for the argument mining task. [79].

8.4 Vocabularies, Semantic Web and Linked Data Based Knowledge Representation and Artificial Intelligence Formalisms on the Web

8.4.1 Semantic Web for Life Sciences

Participants: Pierre Monnin.

Life sciences produce and consume vast amounts of scientific data. The graph-structured nature of these data naturally leads to data-driven research efforts leveraging Semantic Web and Knowledge Graph technologies. Among such usages, knowledge graph construction and management is a well established topic. One subtask lies in matching similar or related units across datasets to identify possible overlaps. In this direction, this year again, we proposed the track “Pharmacogenomics” in the international challenge “Ontology Aligement Evaluation Initiative”. This track focuses on the matching of pharmacogenomic knowledge units, which are n -ary tuples involving components of three distinct types (drugs, genetic factors, and phenotypes). This year again, none of the approaches participating in the 2024 campaign were able to produce alignments. These results highlight once more the interest in considering domain-specific problems, bringing additional challenges to the field of ontology matching. Given the inadequacy of ontology matching systems to produce valid alignments, such challenges are currently unaddressed and require to design new methods or enrich existing ones.

8.4.2 W3C Data activity and AC Rep

Participants: Rémi Ceres, Pierre-Antoine Champin, Fabien Gandon, Franck Michel, Olivier Corby.

Semantic Web technologies are based on a set of standards developed by the **World Wide Web consortium (W3C)**. Participation in these standardization groups gives to researcher the opportunity to promote their results towards a broad audience, and to keep in touch with an international community of experts. Wimmics has a long history of being involved in W3C groups.

As W3C fellow, Pierre-Antoine Champin also works within the W3C team to support Semantic Web related working groups and promote the emergence of new ones, to ensure the necessary evolutions of these technologies. In 2024, the new **Linked Web Storage** Working Group was chartered, to standardize the Solid protocol. The **Solid project** was started by Tim Berners-Lee, inventor of the Web, and builds on Semantic Web standards to promote the (re-)decentralization of the Web. Solid has been a research topic for Wimmics in the past years, including in the collaboration with Startin'Blox (see Section 9.1). The **RDF-star** Working Group is pursuing its efforts to publish the new version of RDF and SPARQL, extending them with the ability to make statements about statements. A new **Data Shapes** Working Group was created in December 2024 to adapt SHACL to those changes in RDF. We intend to reflect those changes into Corese (see Section 7.1.4); in fact, Corese already implements an experimental version of RDF-star.

Finally, Fabien Gandon remains the W3C AC Rep for Inria representing institute in all standardization processes and W3C meetings (annual W3C TPAC conference and W3C AC Meeting).

8.4.3 AutomaZoo: Automatic annotation of an ancient Zoological Corpus

Participants: Arnaud Barbe, Molka Dhouib, Catherine Faron, Franck Michel.

This activity is part of the HISINUM project funded by the Academy of Excellence 5 of UCAJedi and related to the Zoomathia international research network which aims to study the constitution and transmission of zoological knowledge from Antiquity to the Middle Ages. The aim is to produce a corpus of textual resources semantically annotated by a graph of ancient zoological knowledge, respecting semantic web standards, interoperable and published on the open data web.

In previous steps, we built the ZooKG-Pliny knowledge graph that represents the manual annotations of the ancient zoological text *Naturalis Historia* (Pliny the Elder). The semantic annotation is done using concepts from the thesaurus TheZoo [120]. ZooKG-Pliny allows the integration and the interrogation of relevant knowledge in order to support epistemologists, historians and philologists in their analysis of these texts and knowledge transmission through them [112]. We also developed approaches to automatically classify paragraphs into one or more macro collections of concepts (i.e. "Places", "Anthroponym", etc.) from TheZoo [122].

In 2024, we continued the development of the text processing pipeline for a corpus of texts on animals compiled within the framework of the Zoomathia GDRI. The web interface allows researchers to explore the corpus via a search by concept, explore a selected work while visualizing the concepts annotating each of its parts, and visualize the results of queries implementing competency questions on a selected work from the corpus (7.1.31).

8.4.4 A Unified approach to publish semantic annotations of agricultural documents as knowledge graphs

Participants: Nadia Yacoubi Ayadi, Catherine Faron, Franck Michel.

This work was carried out as part of the D2KAB project (Data to Knowledge in Agriculture and Biodiversity), which aims to develop semantic web-based tools to describe and make agronomical data actionable and accessible following the FAIR principles. We focus on constructing domain-specific Knowledge Graphs (KGs) from textual data sources, using Natural Language Processing (NLP) techniques to extract and structure relevant entities. Our approach is based on the formalization of a semantic data model using common linked open vocabularies such as the Web Annotation Ontology (OA) and the Provenance Ontology (PROV). The model was developed by formulating motivating scenarios and

competency questions from domain experts. This model has been used to construct three different KGs from three distinct corpora: PubMed scientific publications on wheat and rice genetics and phenotyping, and French agricultural alert bulletins. The named entities to be recognized include genes, phenotypes, traits, genetic markers, taxa and phenological stages normalized using semantic resources such as the Wheat Trait and Phenotype Ontology (WTO), the French Crop Usage (FCU) thesaurus and the Plant Phenological Description Ontology (PPDO). Named entities were extracted using different NLP approaches and tools. The relevance of the semantic model was validated by implementing experts questions as SPARQL queries to be answered on the constructed RDF knowledge graphs. Our work demonstrates how domain-specific vocabularies and systematic querying of KGs can reveal hidden interactions and support agronomists in navigating vast amounts of data. The resources and transformation pipelines developed are publicly available in Git repositories [41]).

Leveraging this work, we also contributed to the alignment of descriptions of plants with distinct viewpoints [83].

8.4.5 Semantic Web for Biodiversity

Participants: Franck Michel, Catherine Faron.

This activity addresses the challenges of exploiting knowledge representation and semantic Web technologies to enable data sharing and integration in the biodiversity area.

Through the [GBIF CFP Capacity Enhancement Support Programme](#), we have collaborated with the Museum of Natural History of Paris and GBIF Spain in a joint project to implement the representation of species life traits using the Plinian core vocabulary [121]. This entailed the development of a new software (7.1.24) to translate the XSD representation of Plinian Core into an OWL ontology.

8.4.6 Ontology engineering: tooling and methodology

Participants: Fabien Gandon, Nicolas Robert.

We contributed to the Agile and Continuous Integration for Modular Ontologies and Vocabularies (ACIMOV) [63] ontology engineering methodology for developing ontologies and vocabularies. ACIMOV extends the SAMOD agile methodology to (1) ensure alignment to selected reference ontologies; (2) plan module development based on dependencies; (3) define ontology modules that can be specialized for specific domains; (4) empower active collaboration among ontology engineers and domain experts; (5) enable application developers to select views of the ontology for their specific domain and use case. ACIMOV adopts the standard git-based approach for coding, leveraging agility and DevOps principles. It was implemented in OLIVAW using the collaborative software development platforms Github tooling with continuous integration and continuous deployment workflows (CI/CD workflows) that run syntactic and semantic checks on the repository, specialize modules, generate and publish the ontology documentation [111].

8.5 Analyzing and Reasoning on Heterogeneous Semantic Graphs

8.5.1 Corese Semantic Web Factory

Participants: Rémi Ceres, Fabien Gandon, Olivier Corby.

Corese [113], an open-source Semantic Web platform, implements W3C languages such as RDF, RDFS, OWL RL, SHACL, SPARQL, and extensions including SPARQL Function, SPARQL Transformation, and SPARQL Rule.

In the enhancement of Corese's distribution, two new interfaces, Corese-GUI and Corese-Command, were launched on Flathub. Additionally, a one-click installation script for Corese-Command is now available for Linux and MacOS.

The documentation of Corese has been fully updated and is accessible ⁵.

The new interface, Corese-Command, supplements existing ones such as Corese-Library, Corese-GUI, Corese-Server, and Corese-Python. Corese-Command, evolving from the previous Corese-CLI, enables command-line usage of Corese. It encompasses subcommands for converting RDF file formats, running SPARQL queries, performing SHACL validation on RDF datasets, and executing SPARQL queries on remote endpoints. Improvements in file loading now allow handling of local files, URLs, or directories.

All interfaces have been unified to support Corese configuration files in properties format.

Enhancements include bug fixes in Corese-Python, addition of Markdown result format for SPARQL, and N-Quads RDF serialization.

Relevant websites include the Corese project site at [Corese Web site](#) and the GitHub repository at [Corese github URL](#).

8.5.2 Explanatory Argumentation in Natural Language for Correct and Incorrect Medical Diagnoses

Participants: Benjamin Molinet, Elena Cabrio, Serena Villata.

In the context of the ANTIDOTE project, we investigated the generation of natural language argument-based explanations in medicine [40]. A huge amount of research is carried out nowadays in Artificial Intelligence to propose automated ways to analyze medical data with the aim to support doctors in delivering medical diagnoses. However, a main issue of these approaches is the lack of transparency and interpretability of the achieved results, making it hard to employ such methods for educational purposes. It is therefore necessary to develop new frameworks to enhance explainability in these solutions. We present a novel full pipeline to generate automatically natural language explanations for medical diagnoses. The proposed solution starts from a clinical case description associated with a list of correct and incorrect diagnoses and, through the extraction of the relevant symptoms and findings, enriches the information contained in the description with verified medical knowledge from an ontology. Finally, the system returns a pattern-based explanation in natural language which elucidates why the correct (incorrect) diagnosis is the correct (incorrect) one.

8.5.3 ACTA Module Upgrade

Participants: Cristian Cardellino, Theo Alkibiades Collias.

For the ANTIDOTE project, we were in charge of a major refactoring of the original ACTA module ⁶, enabling the latest Large Language Models (LLM) models from Hugging Face. The new module is a standalone Python Library⁷ available under Apache License. We are also in the process of refactoring the code for the web application of ACTA to better integrate it with the new modules and writing a code easier to maintain. We have also integrated the new Medical T5, resulting from a collaboration within ANTIDOTE, to the ACTA demo [61]. This new version has been published as demo paper at ECAI 2024 [55].

8.5.4 RDF Mining

Participants: Ali Ballout, Catherine Faron, Pierre Monnin, Rémi Felin, Andrea Tettamanzi.

⁵Documentation of Corese

⁶Acta module web site

⁷Acta library on Github

The Shapes Constraint Language (SHACL) is a W3C recommendation which allows to represent constraints in RDF, and validate RDF data graphs against these constraints. Acquiring representative and meaningful SHACL constraints from complex and large RDF data graphs is very challenging and tedious. We proposed an approach for the automatic generation of these constraints. It relies on the probabilistic SHACL validation framework to consider the inherent errors in RDF data that we developed in 2023 [114]. It is based on grammatical evolution (GE) for extracting representative SHACL constraints by mining an RDF data graph [60, 84]. We designed RDFminer (7.1.25), an open source Web application to automatically discover SHACL shapes through an evolutionary process. It takes an RDF data graph as input, from which shapes are mined and assessed using a probabilistic validation framework. The user can interact with RDFminer through a dashboard where they can launch and monitor the mining of shapes, and analyze the results in real time. [104]. The results of this research line were defended in the PhD thesis of Rémi Felin [13].

The task of evaluating the fitness of a candidate axiom against known facts or data is known as *candidate axiom scoring*. Being able to accurately score candidate axioms is a prerequisite for automatic schema or ontology induction, but can also be useful for ontology and/or knowledge graph validation. Accurate axiom scoring heuristics are often heavy to compute, which is a big problem if one wants to exploit them in iterative search methods like level-wise generate-and-test or evolutionary algorithms, where large numbers of candidate axioms need to be scored. We have tackled the challenge of learning a predictive model as a surrogate to reasoning, that predicts the acceptability of candidate class axioms, that is fast to execute yet accurate enough to be used in such settings. For this purpose, we have leveraged a semantic similarity measure extracted from the subsumption hierarchy of an ontology and we have proven that our proposed method is able to learn the acceptability labels of candidate OWL class axioms with high accuracy and that it can do so for multiple types of OWL class axioms [48]. The results were defended in the PhD thesis of Ali Ballout [94].

8.5.5 Capturing Geospatial Knowledge from Real-Estate Advertisements

Participants: Lucie Cadorel, Andrea Tettamanzi.

In the framework of a CIFRE thesis with Septeo PropTech, we have proposed a workflow to extract geographic and spatial entities based on a BiLSTM-CRF architecture (Bidirectional Long Short-Term Memory Conditional Random Field) with a concatenation of several text representations and to extract spatial relations, to build a structured Geospatial knowledge base. This pipeline has been applied to the case of French housing advertisements, which generally provide information about a property's location and neighborhood. Our results show that the workflow tackles French language and the variability and irregularity of housing advertisements, generalizes Geoparsing to all geographic and spatial terms, and successfully retrieves most of the relationships between entities from the text.

Text representations are widely used in NLP tasks such as text classification. Very powerful models have emerged and been trained on huge corpora for different languages. However, most of the pre-trained models are domain-agnostic and fail on domain-specific data. We performed a comparison of different text representations applied to French Real Estate classified advertisements through several text classification tasks to retrieve some key attributes of a property. Our results demonstrate the limitations of pre-trained models on domain-specific data and small corpora, but also the strength of text representation, in general, to capture underlying knowledge about language and stylistic specificities.

The results of this research line were defended in the PhD thesis of Lucie Cadorel [95].

8.5.6 Outlier Detection in MET data using Subspace Outlier Detection Method

Participants: Dupuy Rony Charles, Andrea Tettamanzi.

In plant breeding, Multi-Environment Field Trials (MET) are commonly used to evaluate genotypes for multiple traits and to estimate their genetic breeding value using Genomic Prediction (GP). The

occurrence of outliers in MET is common and is known to have a negative impact on the accuracy of the GP. Therefore, identification of outliers in MET prior to GP analysis can lead to better results. However, Outlier Detection (OD) in MET is often overlooked. Indeed, MET give rise to different level of residuals which favor the presence of swamping and masking effects, where ideal sample points may be portrayed as outliers instead of the true ones. Consequently, without a sensitive and robust outlier detection algorithm, OD can be a waste of time and potentially degrades the accuracy prediction of the GP, especially when the data set is not huge. In this study, we compared various robust outlier methods from different approaches to determine which one is most suitable for identifying MET anomalies. Each method has been tested on eleven real-world MET data sets. Results are validated by injecting a proportion of artificial outliers in each set. The Subspace Outlier Detection Method stands out as the most promising among the tested methods [56].

8.5.7 Estimating User's Knowledge Gain in Search-as-Learning Using Knowledge Graphs

Participants: Andrea Tettamanzi.

In the context of search as learning, users engage in search sessions to fill their information gaps and achieve their learning goals. Tracking the user's state of knowledge is therefore essential for estimating how close they are to achieve these learning goals. In this respect, we extend a recently proposed approach that uses the recognition of entities present in the text to track the user's knowledge. Our approach introduces a more complete representation by considering both the entities and their relations. More precisely, we represent both the user's knowledge and the user's learning goals (or target knowledge) as knowledge graphs. We show that the proposed representation captures a complementary aspect of knowledge, thus helping to improve the user knowledge gain estimation when used in combination with other representations [69].

8.5.8 ISSA: semantic indexing of scientific articles and advanced services

Participants: Anna Bobasheva, Catherine Faron, Aline Menin, Franck Michel, Marco Winckler.

The *ISSA 2* project, started in 2023, was completed by the end of 2024. It proposes a generic method to explore the data available in an open archive in order to 1) extract new knowledge, 2) exploit this knowledge with a bibliometric objective and 3) propose services to researchers and documentalists, in particular in terms of bibliometrics and information retrieval. It relies on and extends the outcomes of the ISSA project, in particular the semantic index of the publications of an open archive (metadata, descriptors and named entities mentioned in the text, linked to standard knowledge bases). The proposed methods exploit data mining techniques as well as techniques for the construction, publication and exploitation of knowledge graphs. In addition to Agritrop, CIRAD's open archive that served as a use case in ISSA, ISSA 2 considers the HAL instance of the UR EuroMov Digital Health in Motion.

In 2024, we extended the indexing pipeline (see Section 7.1.14, and [110]) to gather additional data about the disciplines and Sustainable Development Goals of each article, and compute the Rao-Stirling diversity index. We also enriched existing interfaces to visualize these new data.

The pipeline and search interface are DOI-identified and available under an open license on public repositories. The knowledge graph produced by the pipeline for the Agritrop archive is also made public as a downloadable dump (DOI: [10.5281/zenodo.10381606](https://doi.org/10.5281/zenodo.10381606)) and through a public SPARQL endpoint.

8.5.9 An Open Platform for Quality Measures in a Linked Data Index

Participants: Pierre Maillot, Olivier Corby, Catherine Faron, Fabien Gandon, Franck Michel.

There is a great diversity of RDF datasets publicly available on the web. Choosing among them requires assessing their “fitness for use” for a particular use case, and thus, finding the right quality measures and evaluating data sources according to them. However, this is not an easy task due to the large number of possible quality measures, and the multiplicity of implementation and assessment platforms. Therefore, there is a need for a common way to define measures and evaluate RDF datasets, using open standards and tools.

Developed in the context of the ANR DeKaloG, IndeGx is a SPARQL-based framework to design indexes of Knowledge Graphs declaratively [118]. We extended it to support more advanced data quality measures. We demonstrated our approach by reproducing two existing measures, showing how one can formalize and add measures using such an open declarative framework. This work was presented at the Web Conference 2024 [66].

8.5.10 Learning Pattern-Based Extractors from Natural Language and Knowledge Graphs: Applying Large Language Models to Wikipedia and Linked Open Data

Participants: Célian Ringwald, Fabien Gandon, Catherine Faron, Franck Michel, Hanna Abi Akl.

Whether automatically extracted from structured elements of articles or manually populated, the open and linked data published in DBpedia, and Wikidata offer rich and structured complementary views of the textual descriptions found in Wikipedia. However, the unstructured text of Wikipedia articles contains a lot of information that is still missing in DBpedia and Wikidata. Extracting them would be interesting to improve the coverage and quality of these knowledge graphs (KG) and this would have an important impact on all downstream tasks.

This work proposes to exploit the dual bases formed from Wikipedia pages and Linked Open Data (LOD) bases covering the same subjects in natural language and in RDF, to produce RDF extractors targeting specific RDF patterns and tuned for a given language. Therefore, the main research question is: Can we learn efficient customized extractors targeting specific RDF patterns from the dual base formed by Wikipedia on one hand, and DBpedia and Wikidata on the other hand?

The landscape of the research field drawn at the intersection of language models and knowledge graphs is very dynamic and quickly evolving. For this reason, as the first step of this work, we designed an extended systematic review of the latest NLP approaches to KG extraction.

In a second step, we started the design a first dataset focused on datatype properties. We restricted the selection of our training to facts respecting a given SHACL shape and information that could be found in the Wikipedia abstract. Then, to learn how to extract relations with datatype properties from natural language, we exploited pre-trained encoder-decoder models, and more precisely BART (denoising autoencoder sequence-to-sequence model). We explored several aspects of the task formulation that could impact the generation of triples in this context: the size of the model, the size of the learning sample needed to learn a given SHACL pattern, and the syntax of the triples [109, 108].

We continued the work by questioning the impact of the syntax chosen for representing the generated output by benchmarking 12 variations of RDF syntaxes, but also by comparing two small language models (T5 and BART), and demonstrated the performances of a light Turtle syntax [76].

Future works include the systematic extension of the pattern-based relation extraction, as well as the extraction of object properties.

8.5.11 AI agent to convert natural language questions into SPARQL queries

Participants: Franck Michel, Fabien Gandon.

An experimental knowledge graph (KG) driven framework (10.26434/chemrxiv-2023-sljbt) was recently introduced to facilitate the integration of heterogeneous data types, encompassing both experimental data (mass spectrometry annotation, results from biological screening and fractionation) as

well as meta-data available on the Web (such as taxonomies and metabolite databases). Although this KG efficiently encapsulates the different data structures and semantic relationships, retrieving specific information through structured or visual queries or even programmatically, is not trivial.

In the collaborative project MetaboT (formerly KG-Bot), we designed and implemented an AI agent that can convert natural language questions into SPARQL queries and programmatic data-mining tasks, and generate adapted visualization. By leveraging the potential of emerging Large Language Models (LLMs) to understand semantic relationships encapsulated in KGs and mentioned in the questions, the agent autonomously iterates to construct a SPARQL query of any submitted natural language question. After retrieving the necessary information from the KG, the agent provides a preliminary interpretation of the results in natural language, along with relevant visualizations and statistics [123].

Following this work, we have submitted an ANR project proposal called MetaboLinkAI. This project was accepted and shall start in 2025.

8.5.12 Hybridizing machine learning and knowledge graphs: injection of relation signatures

Participants: Pierre Monnin.

Knowledge graphs (KGs) are nowadays largely adopted, representing a successful paradigm of how symbolic and transparent AI can scale on the World Wide Web. However, they are generally tackled by Machine Learning (ML) and mostly numeric based methods such as graph embedding models (KGEMs) and deep neural networks (DNNs). The latter methods have been proved efficient but lack major characteristics such as interpretability and explainability. Conversely, these characteristics are intrinsically supported by symbolic AI methods and artefacts, thus motivating a research effort to hybridize machine learning and knowledge graphs.

Towards such an hybridization, we investigated the improvement of KGEMs with symbolic knowledge for the task of link prediction which aims at predicting the missing tail of a triple $(h, r, ?)$ or the missing head of a triple $(?, r, t)$. To train such methods, batches of positive and negative triples are considered in loss functions. However, different kinds of negative triples exist: considering signatures of relations (domain and range), some negative triples may be semantically valid (e.g., $(\text{PierreMonnin}, \text{researchTeam}, \text{Wimmics})$), while others may be semantically invalid (e.g., $(\text{PierreMonnin}, \text{researchTeam}, \text{Mediterranean Sea})$). We enriched the three main loss functions for link prediction such that all kinds of negatives are sampled but treated differently based on their semantic validity [64, 85]. This enrichment led to improved semantic correctness of KGEMs as measured by the Sem@K metric, but also better ranking of gold truth entities as measured by the MRR and Hits@10 metrics. Such results highlight the interest of considering symbolic knowledge in KGEMs.

8.5.13 Enriching benchmark datasets by generating synthetic ontologies and knowledge graphs

Participants: Pierre Monnin.

In tasks of the lifecycle of knowledge graphs (e.g., link prediction, entity classification), only a few KGs established themselves as standard benchmarks for evaluating approaches. However, recent works outline that relying on a limited collection of datasets is not sufficient to assess the generalization capability of an approach. To remedy the aforementioned issues, we released PyGraft, a Python library to generate synthetic ontologies and knowledge graphs by configuring the logical (RDFS and OWL) constructions to use [65, 86]. By providing a way of generating both a schema and KG in a single pipeline, this library will support the development a more diverse array of KGs for benchmarking novel approaches. Especially, PyGraft will support the development and evaluation of new neuro-symbolic AI approaches by allowing to integrate and evaluate the contribution of different types of knowledge constructs, thereby going beyond what is currently possible with the limited collection of available benchmarks.

8.5.14 Reusing Wikidata to build new knowledge graphs: the KGPrune platform

Participants: Pierre Monnin.

More and more knowledge graphs are publicly published and accessible on the Web of data, covering a widening array of domains. This allows their reuse in other downstream applications or in the construction of other knowledge graphs. However, not all represented knowledge is useful or pertaining in such cases. This is particularly the case for general large-scale knowledge graphs such as Wikidata. Additionally, the sheer size of such knowledge graphs entails scalability issues. These two aspects ask for efficient methods to extract subgraphs of interest from existing knowledge graphs. To this aim, we introduce KGPrune, an API and Web Application that, given seed entities of interest and properties to traverse, extracts their neighboring subgraphs from Wikidata [107, 43]. To avoid topical drift, KGPrune relies on a frugal pruning algorithm based on analogical reasoning to only keep relevant neighbors while pruning irrelevant ones.

8.5.15 Formal Argumentation: How can we understand and deal with implicit arguments?

Participants: Victor David.

Linking Argument Mining and Formal Argumentation to better understand implicit arguments:

Argument mining is a natural language processing technology aimed at identifying arguments in texts. Furthermore, the approach is being developed to identify the premises and claims of those arguments, and to identify the relationships between arguments including support and attack relationships. In the paper [100], we assume that an argument map contains the premises and claims of arguments, and support and attack relationships between them, that have been identified by argument mining. So from a piece of text, we assume an argument map is obtained automatically by natural language processing. However, to understand and to automatically analyze that argument map, it would be desirable to instantiate that argument map with logical arguments. Once we have the logical representation of the arguments in an argument map, we can use automated reasoning to analyze the argumentation (e.g. check consistency of premises, check validity of claims, and check the labelling on each arc corresponds with the logical arguments). We address this need by using classical logic for representing the explicit information in the text, and using default logic for representing the implicit information in the text. In order to investigate our proposal, we consider some specific options for instantiation.

Formal Argumentation to better evaluate the quality of decoding implicit arguments:

An argument can be seen as a pair consisting of a set of premises and a claim supported by them. Arguments used by humans are often enthymemes, i.e., some premises are implicit. To better understand, evaluate, and compare enthymemes, it is essential to decode them, i.e., to find the missing premisses. Many enthymeme decodings are possible. We need to distinguish between reasonable decodings and unreasonable ones. However, there is currently no research in the literature on “How to evaluate decodings?”. To pave the way and achieve this goal, we introduce seven criteria related to decoding, based on different research areas. Then, we introduce the notion of criterion measure, the objective of which is to evaluate a decoding with regard to a certain criterion. Since such measures need to be validated, we introduce several desirable properties for them, called axioms. Another main contribution of the paper [99] is the construction of certain criterion measures that are validated by our axioms. Such measures can be used to identify the best enthymemes decodings.

8.5.16 How to reason with probabilistic and temporal information?

Participants: Victor David, Pierre Monnin.

A new and faster algorithm for reasoning on probabilistic argumentation graphs:

The paper [49] presents fast and exact methods for computing the probability of an argument's acceptance using Dung's semantics in the Constellation paradigm of Abstract Argumentation. For (directed) Singly-Connected Graphs (SCGs), the problem can now be solved in linearithmic time instead of being exponential in the number of attacks, as reported in the literature. Moreover, in the more general case of Directed Acyclic Graphs (DAGs), we provide an algorithm whose time complexity is linearithmic in the product of the out-degree of dependent arguments, i.e., arguments reaching the argument considered for acceptance through multiple paths in the graph. We theoretically show that this complexity is lower than the lower-bound of the (exact) Constellation method, which is also supported by empirical results. We also compare our approach on DAGs with the (approximate) Monte-Carlo method, which is stopped when our approach obtains the exact results. Within this time constraint, Monte-Carlo still outputs significant errors, underlying the fast computation of our approach.

A new model able to represent and reason about probabilistically uncertain arguments over a time interval:

The study of Dung-style Argumentation Frameworks in recent years has focused on incorporating time. For example, availability intervals have been added to arguments and relations, resulting in different outputs of Dung semantics over time. The paper [36] examines the probability distribution of arguments over time intervals. Using this temporal probabilistic model, the study explores how these frameworks can be transformed into a probabilistic argumentation according to the constellation approach and how they can be interpreted within the epistemic approach. The epistemic approach relies on the notion of defeat to select significant conflicts based on probability distributions. The study also introduces the temporal acceptability of arguments based on the concept of defence, allowing for more precise results over time. Finally, the models (constellation and epistemic) are extended to account for events that have a duration, i.e., that can occur for several consecutive instants of time.

A Knowledge-Graphs-based approach for extracting the maximal subset of coherent information (i.e., the MAP inference task) from contradictory, uncertain and temporal data:

Reasoning on inconsistent and uncertain data is challenging, especially for Knowledge-Graphs (KG) to abide temporal consistency. Our goal is to enhance inference with more general time interval semantics that specify their validity, as regularly found in historical sciences. We propose a new Temporal Markov Logic Networks (TMLN) model which extends the Markov Logic Networks (MLN) model with uncertain temporal facts and rules. Total and partial temporal (in)consistency relations between sets of temporal formulae are examined. We then propose a new Temporal Parametric Semantics (TPS) which allows combining several sub-functions leading to different assessment strategies. Finally, in [89] we present the NeoMaPy tool, to compute the MAP inference on MLNs and TMLNs with several TPS. We compare our performances with state-of-the-art inference tools and exhibit faster and higher quality results.

9 Bilateral contracts and grants with industry

9.1 Bilateral contracts with industry

CIFRE Contract with Doriane

Participants: Andrea Tettamanzi, Rony Dupuy Charles.

Partner: Doriane.

This collaborative contract for the supervision of a CIFRE doctoral scholarship, relevant to the PhD of Rony-Dupuy Charles, is part of Doriane's Fluidity Project (Generalized Experiment Management),

the feasibility phase of which has been approved by the Terralia cluster and financed by the Région Sud-Provence Alpes Côte d'Azur and BPI France in March 2019. The objective of the thesis is to develop machine learning methods for the field of agro-vegetation-environment. To do so, this research work will take into account and address the specificities of the problem, i.e. data with mainly numerical characteristics, scalability of the study object, small data, availability of codified background knowledge, need to take into account the economic stakes of decisions, etc. To enable the exploitation of ontological resources, the combination of symbolic and connective approaches will be studied, among others. Such resources can be used, on the one hand, to enrich the available datasets and, on the other hand, to restrict the search space of predictive models and better target learning methods.

The PhD student developed original methods for the integration of background knowledge in the process of building predictive models and for the explicit consideration of uncertainty in the field of agro-plant environment.

CIFRE Contract with Kinaxia

Participants: Andrea Tettamanzi, Lucie Cadorel.

Partner: Kinaxia.

This thesis project is part of a collaboration with Kinaxia that began in 2017 with the Incertimmo project. The main theme of this project was the consideration of uncertainty for a spatial modeling of real estate values in the city. It involved the computer scientists of the Laboratory and the geographers of the **ESPACE Laboratory**. It allowed the development of an innovative methodological protocol to create a mapping of real estate values in the city, integrating fine-grained spatiality (the street section), a rigorous treatment of the uncertainty of knowledge, and the fusion of multi-source (with varying degrees of reliability) and multi-scale (parcel, street, neighbourhood) data.

This protocol was applied to the Nice-Côte d'Azur metropolitan area case study, serving as a test bed for application to other metropolitan areas.

The objective of this thesis, which was carried out by Lucie Cadorel under the supervision of Andrea Tettamanzi, was, on the one hand, to study and adapt the application of methods for extracting knowledge from texts (or text mining) to the specific case of real estate ads written in French, before extending them to other languages, and, on the other hand, to develop a methodological framework that makes it possible to detect, explicitly qualify, quantify and, if possible, reduce the uncertainty of the extracted information, in order to make it possible to use it in a processing chain that is finalized for recommendation or decision making, while guaranteeing the reliability of the results.

Plan de Relance with Startin'Blox

Participants: Pierre-Antoine Champin, Fabien Gandon, Maxime Lecoq.

Partner: **Startin'Blox**.

The subject of this project is to investigate possible solutions to build on top of the Solid architecture capabilities to discover services and access distributed datasets. This would rely on standardized search and filtering capabilities for the Solid PODs, as well as on traversal or federated SPARQL query solving approaches to design a pilot architecture. We also intend to address performance issues via caching or indexing strategies in order to allow a deployment of the Solid ecosystem on a web scale.

An outcome of this collaboration is a vocabulary for describing indexes in the Solid ecosystem. This vocabulary will serve as a starting point for a work item in the **W3C Solid Community Group**.

CP4SC Project - sub contract

Participants: Fabien Gandon, Rémi Ceres.

Partner: Atos.

Initiated in January 2023, the CP4SC project is a collaborative effort with Atos that focuses on developing a Cloud Platform for Smart Cities. This innovative platform is designed to integrate and analyze data from a wide range of city services and smart devices. As a single, secure access point for data management and dissemination, it promotes innovation and interdepartmental cooperation in urban areas. The project targets sustainable urban development by addressing three key areas: optimizing resource consumption in various living environments through energy management; reducing greenhouse gas emissions via mobility management and alternative transportation solutions; and enhancing environmental conservation and public health through strategic management and observation of the Earth's environment. In that context, Atos has a contract with Wimmics to study the use of semantic technologies and CORESE in managing the data of this Smart City scenario.

10 Partnerships and cooperations

10.1 International research visitors

10.1.1 Visits of international scientists

Paolo Buono

Status: (researcher)

Institution of origin: Università degli Studi di Bari Aldo Moro

Country: Italy

Dates: July 19, 2024 - September 7, 2024

Context of the visit: This visit is part of an effort to enforce the collaboration between the University of Bari and INRIA/Université Côte d'Azur on the topics around user interaction and Semantic Web. The visit of professor Buono was funded by I3S.

Mobility program/type of mobility: (research stay)

Marieke van Erp

Status: (researcher)

Institution of origin: KNAW Humanities Cluster

Country: The Netherlands

Dates: September 9, 2024 - September 27, 2024

Context of the visit: This visit aims to initiate collaborations in Digital Humanities.

Mobility program/type of mobility: (research stay)

10.1.2 Visits to international teams

Research stays abroad

Pierre Monnin

Visited institution: Università degli Studi di Bari Aldo Moro

Country: Italy

Dates: January 15–February 2, 2024

Context of the visit: bilateral collaboration on the development of neuro-symbolic approaches for the discovery and injection of symbolic knowledge in knowledge graph embedding models

Mobility program/type of mobility: research stay

Visited institution: INESC-ID, Instituto Superior Técnico, Universidade de Lisboa

Country: Portugal

Dates: November 14–28, 2024

Context of the visit: collaboration on analogical reasoning, in the framework of the AT2TA ANR project

Mobility program/type of mobility: research stay

Marco Winckler

Visited institution: Università degli Studi di Bari Aldo Moro

Country: Italy

Dates: October 23, 2023 - January 15, 2024

Context of the visit: This visit is part of an effort to enforce the collaboration between the University of Bari and INRIA/Université Côte d'Azur on the topics around user interaction and Semantic Web. The visit was funded by I3S.

Mobility program/type of mobility: research stay including teaching activities and students supervision.

Anaïs Ollagnier

Visited institution: Università degli Studi di Torino

Country: Italy

Dates: June 16, 2024 - July 13, 2024

Context of the visit: This visit is part of an effort to strengthen the collaboration between the University of Torino and INRIA/Université Côte d'Azur on topics related to natural language understanding, focusing on pragmatics and perspectivism. The visit was funded by I3S as part of the AAP AAPI initiative.

Mobility program/type of mobility: research stay.

10.2 European initiatives

10.2.1 H2020 projects

AI4Media

Title: AI4Media

Duration: 2020 - 2024

Coordinator: The Centre for Research and Technology Hellas (CERTH)

Partners:

see consortium

Inria contact: through 3IA

Summary: AI4Media is a 4-year-long project. Funded under the European Union's Horizon 2020 research and innovation programme, the project aspires to become a Centre of Excellence engaging a wide network of researchers across Europe and beyond, focusing on delivering the next generation of core AI advances and training to serve the Media sector, while ensuring that the European values of ethical and trustworthy AI are embedded in future AI deployments. AI4Media is composed of 30 leading partners in the areas of AI and media (9 Universities, 9 Research Centres, 12 industrial organizations) and a large pool of associate members, that will establish the networking infrastructure to bring together the currently fragmented European AI landscape in the field of media, and foster deeper and long-running interactions between academia and industry.

ORBIS

Title: Augmenting Participation, Co-creation, Trust and Transparency in Deliberative Democracy (ORBIS)

Duration: 2023 - 2026

Coordinator: Politecnico di Milano (Italy)

Partners:

[see consortium](#)

Inria contact: Elena Cabrio, Serena Villata

Summary: ORBIS responds to the profound lack of dialogue between citizenship and policy making institutions by providing a theoretically sound and highly pragmatic socio-technical solution to enable the transition to a more inclusive, transparent and trustful Deliberative Democracy in Europe.

10.2.2 Other european programs/initiatives

HyperAgents - SNSF/ANR project

Title: HyperAgents

Duration: 2020 - 2024

Coordinator: Olivier Boissier, MINES Saint-Étienne

Partners:

- MINES Saint-Étienne (FR)
- INRIA (FR)
- Univ. of St. Gallen (HSG, Switzerland)

Inria contact: Fabien Gandon

Summary: The HyperAgents project, Hypermedia Communities of People and Autonomous Agents, aims to enable the deployment of world-wide hybrid communities of people and autonomous agents on the Web. For this purpose, HyperAgents defines a new class of multi-agent systems that use hypermedia as a general mechanism for uniform interaction. To undertake this investigation, the project consortium brings together internationally recognized researchers actively contributing to research on autonomous agents and MAS, the Web architecture, Semantic Web, and to the standardization of the Web. Project Web site: [HyperAgents Project](#)

ANTIDOTE - CHIST-ERA project

Title: ANTIDOTE

Duration: 2020 - 2024

Coordinator: Elena Cabrio, Serena Villata

Partners:

- University of the Côte d’Azur (Wimmics Team)
- Fondazione Bruno Kessler (IT)
- University of the Basque Country (ES)
- University of Leuven (Belgium)
- University of Lisbon (PT)

Summary: Providing high quality explanations for AI predictions based on machine learning requires to combine several interrelated aspects, including, among others: selecting a proper level of generality/specificity of the explanation, considering assumptions about the familiarity of the explanation beneficiary with the AI task under consideration, referring to specific elements that have contributed to the decision, making use of additional knowledge (e.g. metadata) which might not be part of the prediction process, selecting appropriate examples, providing evidences supporting negative hypothesis, and the capacity to formulate the explanation in a clearly interpretable, and possibly convincing way. According to the above considerations, ANTIDOTE fosters an integrated vision of explainable AI, where low level characteristics of the deep learning process are combined with higher level schemas proper of the human argumentation capacity. ANTIDOTE will exploit cross-disciplinary competences in three areas, i.e. deep learning, argumentation and interactivity, to support a broader and innovative view of explainable AI. Although we envision a general integrated approach to explainable AI, we will focus on a number of deep learning tasks in the medical domain, where the need for high quality explanations, both to clinicians and to patients, is perhaps more critical than in other domains. Project Web site: [Antidote Project](#)

10.3 National initiatives

ANR D2KAB

Participants: Olivier Corby, Catherine Faron, Fabien Gandon, Franck Michel, Nadia Yacoubi Ayadi.

Partners: LIRMM, INRAE, IRD, ACTA

D2KAB is an ANR project which started in June 2019 until June 2024, led by the LIRMM laboratory (UMR 5506). Its general objective is to create a framework to turn agronomy and biodiversity data into knowledge –semantically described, interoperable, actionable, open– and investigate scientific methods and tools to exploit this knowledge for applications in science and agriculture. Within this project the Wimmics team is contributing to the lifting of heterogeneous dataset related to agronomy coming from the different partners of the project and is responsible to develop a unique entry point with semantic querying and navigation services providing a unified view on the lifted data.

Web site: [D2KAB Project](#)

ANR DeKaloG

Participants: Olivier Corby, Catherine Faron, Fabien Gandon, Pierre Maillot, Franck Michel.

Partners: Université Nantes, INSA Lyon, Inria Center at Université Côte d'Azur

DeKaloG (Decentralized Knowledge Graphs) is an ANR project until June 2024 that aims to: (1) propose a model to provide fair access policies to KGs without quota while ensuring complete answers to any query. Such property is crucial for enabling web automation, i.e. to allow agents or bots to interact with KGs. Preliminary results on web preemption open such perspective, but scalability issues remain; (2) propose models for capturing different levels of transparency, a method to query them efficiently, and especially, techniques to enable web automation of transparency; (3) propose a sustainable index for achieving the findability principle.

DeKaloG Web site

ANR ATTENTION

Participants: Serena Villata, Elena Cabrio, Xiaou Wang, Pierpaolo Goffredo.

The ANR project ATTENTION started in January 2022 with Université Paris 1 Sorbonne, CNRS (Centre Maurice Halbwachs), EURECOM, Buster.Ai. The coordinator of the project is CNRS (Laboratoire I3S) in the person of Serena Villata.

In the ATTENTION project, we propose to address the urgent need of designing intelligent semi-automated ways to generate counter-arguments to fight the spread of disinformation online. The idea is to avoid the undesired effects that come with content moderation, such as overblocking, when dealing with disinformation online, and to directly intervene in the discussion (e.g., Twitter threads) with textual arguments that are meant to counter the fake content as soon as possible, and prevent it from further spreading. A counter-argument is a non-aggressive response that offers feedback through fact-bound arguments, and can be considered as the most effective approach to withstand disinformation. Our approach aims at obtaining high quality counter-arguments while reducing efforts and supporting human fact-checkers in their everyday activities.

ANR CIGAlA

Participants: Serena Villata, Mariana Eugenia Chaves Espinoza, Elena Cabrio, Sofiane Elguendouze.

The ANR ASTRID CIGAlA project (December 2022 - 30 Months) "Controversy and influence in the Ukraine war: a study of argumentation and counter-argumentation through Artificial Intelligence" aims to bring together, in a multi-disciplinary way, the research fields of discourse analysis in the humanities and social sciences, with those of automatic extraction of natural language arguments from text (argument mining) in Artificial Intelligence.

Partners are Ecole de l'Air et de l'Espace (EAE), which is also the coordinator of the project, and CNRS (Laboratoire I3S).

ANR CROQUIS

Participants: Andrea Tettamanzi.

The ANR project CROQUIS (March 2022 - 48 months) with CRIL (Lens) and HSM (Montpellier). The coordinator of the project is Salem Benferhat (CRIL). The local coordinator for Laboratoire I3S is Andrea Tettamanzi. The local unit involves two other members of I3S which are not part of WIMMICS, namely Célia da Costa Pereira and Claude Pasquier. The contribution of Wimmics is focused on addressing the problem of incomplete and uncertain data.

Web site: [CROQUIS Project](#)

ANR AT2TA

Participants: Pierre Monnin.

Partners: Université de Lorraine (LORIA), Inria Paris (HeKA team), Université Paul Sabatier (IRIT), IHU Imagine, Université Côte d'Azur (I3S), Infologic

The ANR project AT2TA started in February 2023 until February 2026. The coordinator of the project is Miguel Couceiro (LORIA, Université de Lorraine). The local coordinator for I3S / Wimmics is Pierre Monnin. The project aims to develop an analogy-based machine learning framework and to demonstrate its usefulness in real case scenarios. Within the project, the Wimmics team is contributing by investigating the potential usages of analogy-based frameworks with and for knowledge graphs, and the associated adequate representation spaces.

Web site: [AT2TA project](#)

ISSA (AAP Collex-Persée)

Participants: Franck Michel, Anna Bobasheva, Olivier Corby, Catherine Faron, Aline Mennin, Marco Winckler.

Partners: CIRAD, Mines d'Alès

The **ISSA project** started in October 2020 and is led by the CIRAD. It aims to set up a framework for the semantic indexing of scientific publications with thematic and geographic keywords from terminological resources. It also intends to demonstrate the interest of this approach by developing innovative search and visualization services capable of exploiting this semantic index. Agritrop, Cirad's open publications archive, serves as a use case and proof of concept throughout the project. In this context, the primary semantic resources are the Agrovoc thesaurus, Wikidata and GeoNames.

Wimmics team is responsible for (1) the generation and publication of the knowledge graph representing the indexed entities, and (2) the development of search/visualization tools intended for researchers and/or information.

CROBORA: Crossing Borders Archives: understanding the circulation of images of Europe (ANR)

Participants: Marco Winckler, Aline Menin.

Coordinator: Matteo Treleani, SicLab, Université Côte d'Azur

The **CROBORA project** (ANR-20-CE38-0002) project led by the Sic.Lab laboratory at EUR CREATES, University of Côte d'Azur, funded by ANR from 2021 to 2024. CROBORA studies the circulation of archive images in the media space. The main hypothesis of the project is that what determines the circulation of archives, thus constituting the visual memory of the European construction, is not a decision that lies solely in the hands of the authors of the reuses (journalists, audiovisual professionals), but rather the consequence of a series of mediations that can be technical (the availability of archives for example), interprofessional (the relationship between archival institutions and the media), cultural (the use of a document for a purpose in one country or another), historical (an archive sequence can change its meaning over time) etc. The general objective of the project is therefore to understand the logics governing the circulation of audiovisual archives. The project aims to respond to the following sub-objectives: (1) to understand which audiovisual fragments are reused in the media to talk about Europe; (2) to build a cartography of frequently used symbolic images; (3) to analyze the representations carried by these images; (4) to understand their trajectory in order to see how they are reshaped diachronically and according to different media, countries and institutions; and finally (5) to identify which mediations determine their readjustments.

Wimmics team is responsible for (i) the development of visualization tools for supporting the exploration of the CROBORA dataset, and (ii) investigating algorithms for optimizing the visual search in the dataset.

CIIAM (Contextual Information Inference for Argument Mining)

Participants: Anaïs Ollagnier.

The CIIAM Project aims to explore the integration of linguistic analysis dimensions related to pragmatics. This approach is particularly innovative, as pragmatics—owing to its complexity and the challenges of formalization—has been largely underexplored from a computational perspective. By utilizing state-of-the-art prompting techniques, the project seeks to address the limitations of existing methods, including challenges such as cultural variability and implicit meaning. The project includes 12 months of funding for a postdoctoral researcher.

Web site: [CIIAM Project](#)

10.4 Regional initiatives

KGBot

Participants: Franck Michel, Yousouf Taghzouti, Fabien Gandon.

Partners: Institute of Chemistry of Nice

The **KGBot project (Knowledge Graph chatBot)** aims to enhance an AI-powered chemistry chatbot prototype designed to improve the accessibility and usability of metabolomics knowledge graphs (KGs). By leveraging mass spectrometry data, the chatbot employs a natural language interface to generate queries (using the SPARQL language), allowing chemists to intuitively explore complex metabolomics knowledge graphs represented in RDF. Key objectives include broadening the chatbot's compatibility with various large language models (LLMs) and KGs, integrating dynamic tools for data extraction and visualization, and enabling extended dialogical interactions to support iterative queries. The project also seeks to enrich user interactions by providing features such as result visualization, hypothesis generation, and analysis recommendations. Building on the interdisciplinary expertise of the project partners, this initiative fosters transdisciplinary collaboration and aims to deliver scalable solutions applicable across multiple domains. Anticipated outcomes include enhanced access to scientific data and the development of a robust open-source framework to support future academic and industrial applications. Funding will be directed towards supporting postdoctoral researchers and student contributions to the project.

11 Dissemination

11.1 Promoting scientific activities

11.1.1 Scientific events: organization

- Organization of The First Workshop on Natural Language Argument-Based Explanations (ArgNLE). Organizers: Elena Cabrio, Rodrigo Agerri, Bernardo Magnini, Marcin Lewinsky, Marie-Francine Moens and Serena Villata. Co-located with the 27th European Conference in Artificial Intelligence. October 20th, 2024.
- Organization of a workshop for the restitution of the results of the ISSA 2 project: “Artificial intelligence at the service of bibliographic research”. Located at Cirad, Montpellier, September 2024. Co-organizer: Franck Michel
- Organization of a workshop on Digital Humanities in the framework of the Hisinum project, September 25th 2024, Maison des Sciences de l’Homme et de la Société Sud-Est. Organizers: Catherine Faron and Arnaud Zucker.

- Organization of the “Journée d’études Extended Reality Research & Creative Center”, third XR2C2 seminar, May 18th 2024, Campus Sophiatech, Sophia Antipolis, France. Co-organizers: Marco Winckler and Aline Menin
- Organization of the “7th IA2 Autumn School” on Artificial Intelligence and Democracy, October 14-18th, 2024, Campus Sophiatech, Sophia Antipolis, France. Co-organizer: Serena Villata, and Victor David

11.1.2 Scientific events: selection

Member of the conference program committees

- Elena Cabrio: member of the Senior Program Committees of the European Conference on Artificial Intelligence (ECAI), of the International Joint Conference on Artificial Intelligence (IJCAI), of the AAI conference, Senior Action Editor of ACL ARR Rolling Review, Senior Area Chair of COLING 2025 (track: Sentiment Analysis, Opinion and Argument Mining).
- Pierre-Antoine Champin: Extended Semantic Web Conference (ESWC) 2024 (resource track), International Semantic Web Conference (ISWC) 2024, The Web Conference 2025
- Catherine Faron: member of Program Committees of European Conference on Artificial Intelligence (ECAI) 2024; International Conference on Knowledge Engineering and Knowledge Management (EKAW) 2024; European Semantic Web Conference (ESWC) 2024, research track PC, resource track senior PC, LLMs for KE special track PC; International Semantic Web Conference (ISWC) 2024; International conference on Semanti Systems (SemanticS 2024); International Conference on Agents and Artificial Intelligence (ICAART) 2025; International Workshop of Semantic Digital Humanities (SemDH2024); International Workshop on Semantic Web and Ontology Design for Cultural Heritage (SWODCH 2024); International Workshop on Natural Scientific Language Processing (NSLP 2024); Journées francophones d’Ingénierie des Connaissances (PFIA-IC) 2024; Rencontres des Jeunes Chercheurs en Intelligence Artificielle (PFIA-RJCIA) 2024; French conference Extraction et Gestion des Connaissances (EGC 2025)
- Fabien Gandon: member of Program Committees of European Conference on Artificial Intelligence (ECAI) 2024; International Conference on Knowledge Engineering and Knowledge Management (EKAW) 2024; European Semantic Web Conference (ESWC) 2024 senior PC, PhD Symposium PC, LLMs for KE special track PC; International Joint Conference on Artificial Intelligence (IJCAI) 2024; International Semantic Web Conference (ISWC) 2024; TheWebConf 2024
- Pierre Monnin: member of the Program Committees of TheWebfConf 2024; European Conference on Artificial Intelligence (ECAI) 2024; International Conference on Knowledge Engineering and Knowledge Management (EKAW) 2024; European Semantic Web Conference (ESWC) 2024 – Resources Track; International Semantic Web Conference (ISWC) 2024 – Posters & Demos Track; European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD) 2024; Rencontre des Jeunes Chercheurs en IA (RJCIA) 2024;
- Serena Villata: member of the Area Chair of the European Conference on Artificial Intelligence (ECAI), of the International Joint Conference on Artificial Intelligence (IJCAI), of the AAI conference, Action Editor of ACL ARR Rolling Review, Area Chair of COLING 2025 (track: Sentiment Analysis, Opinion and Argument Mining).
- Aline Menin: member of the Program Committee of the [IEEE VR](#) and [IAxVR](#) conferences.
- Franck Michel: member of the Program Committee of the [European Conference on AI \(ECAI\) 2024](#), [The Extended Semantic Web Conference \(ESWC\) 2024](#), [The Web Conference 2024](#), the [International Conference on Semantic Systems 2024](#),
- Marco Winckler: member of the Program Committe of the [Advanced Visual Interfaces \(AVI 2024\)](#), [ACM Engineering Interactive Systems \(EICS 2024\)](#), [EuroVIS’2024](#), [The 10th International Working Conference on Human-Centered Software Engineering \(HCSE 2024\)](#), [IEEE VR](#), [Brazilian Symposium](#)

on Human-Computer Interaction (IHC'2024), ACM IMX 2024, The ACM International Conference on Interactive Media Experiences, **S-BPM ONE 2024**, 26th Symposium on Virtual and Augmented Reality (SVR 2024), the 25th International Conference on Web Information Systems Engineering (WISE'2024).

- Victor David: member of the Program Committees of SAC 2025 (40th ACM/SIGAPP Symposium On Applied Computing) on the track KRR (Knowledge Representation and Reasoning); The First Workshop on Natural Language Argument-Based Explanations (ArgNLE) collocation with ECAI 2024; Rencontre des Jeunes Chercheurs en IA (RJCIA) 2024.

Reviewer

- Pierre Monnin: additional reviewer of International Conference on Information and Knowledge Management (CIKM) 2024; additional reviewer of AIMLAI workshop at ECML-PKDD 2024;
- Aline Menin: ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS), EuroVis, International Conference on Interactive Media Experiences (IMX)
- Célian Ringwald: Subreviewer EKAW 2024 Research Track, Subreviewer : ESWC 2024 Research Track, PC member ESWC 2024 LLMs for KE Special Track, PC member EKAW 2024 - **Xtail Workshop**

11.1.3 Journal

Member of the editorial boards

- Elena Cabrio: Journal “Traitement Automatique des Langues (TAL)” and of the Italian Journal of Computational Linguistics (IJCOL).
- Catherine Faron: board member of Transactions on Graph Data and Knowledge; guest editor of a Data & Knowledge Engineering (DKE) special issue on Best papers from EGC 2023
- Pierre Monnin: Transactions on Graph Data and Knowledge
- Serena Villata: Artificial Intelligence and Law, Argument and Computation, Journal of Web Semantics.
- Marco Winckler: Journal of Web Engineering (River Publishers), Interacting with Computers (Oxford Press), Behaviour & Information Technology, PACM Proceedings on Human-Computer Interaction, ACM Sheridam.

Reviewer - reviewing activities

- Elena Cabrio: member of the review committee for the Computational Linguistics journal and the Transactions of the Association for Computational Linguistics (TACL) journal.
- Catherine Faron: reviewer for the Semantic Web Journal (SWJ) and the Journal of Web Semantics (JWS).
- Pierre Monnin: IEEE Transactions on Knowledge and Data Engineering, Data & Knowledge Engineering, AI Communications, Transactions on Graph Data and Knowledge, Semantic Web Journal
- Aline Menin: reviewer for Interacting With Computers (IWC), Visual Information (VISINF), and Computer Animation and Virtual Worlds (CAVW) journals.
- Anaïs Ollagnier: Social Network Analysis and Mining manuscript reviewed October 2024, Language Resources and Evaluation manuscript reviewed October 2024, Journal of Big Data manuscript reviewed April 2024, Cluster Computing manuscript reviewed April 2024. Member of the review committee for Language Resources and Evaluation track.
- Franck Michel: reviewer for the Semantic Web Journal (SWJ).
- Victor David: reviewer for the Journal of Artificial Intelligence Research (JAIR).

11.1.4 Invited talks

- Elena Cabrio:
 - “AM-based decision making in the medical and political domains” Keynote speaker at the 1st International Conference on Recent Advances in Robust Argumentation Machines (RATIO-24). Bielefeld, Germany, June 2024.
 - “Analyse automatique de l’argumentation dans les débats politiques”, Collège de France, in the context of B. Sagot’s chair on “Apprendre les langues aux machines”. January 2024.
- Catherine Faron:
 - Invited panelist at the plenary meeting of the COST action Distributed Knowledge Graph, Viena, September 2024.
- Fabien Gandon
 - “Knowledge Graphs as the Foundation for Interoperable Intelligent Systems ” at the 6th Knowledge Graph and Semantic Web Conference (KGSWC), December 11-13, 2024, University Paris Cité (France), (slides), [105]
 - “Knowledge Graphs: A Pivotal Data Structure for Intelligence”, at workshop ICAIR (Industrial Council on Artificial Intelligence Research), October 2, 2024
- Pierre Monnin
 - Neuro-symbolic approaches for the knowledge graph lifecycle. INESC-ID, Universidade de Lisboa – Seminar of the Language & Speech working group, November 22, 2024, Lisbon, Portugal.
 - Neuro-symbolic approaches for the knowledge graph lifecycle. Università degli Studi di Bari Aldo Moro - Seminar of the PhD Program in Computer Science and Mathematics, January 30, 2024, Bari, Italy.
- Serena Villata
 - Keynote speaker at the 29th International Conference on Natural Language and Information Systems (NLDB-2024), June 2024, Turin, Italy.
 - Keynote speaker at the 10th International Conference on Computational Models of Argument (COMMA-2024), September 2024, Hagen, Germany.
- Anaïs Ollagnier
 - Online Hate Detection in Multi-Party Settings: A Review of Challenges and Solutions. Presented as part of the monthly seminar of the Content-Centered Computing Team. University of Turin, July 4, 2024.
 - Overview of AI: Methodological Foundations and Developments. Presented as part of the 2024 Summer School - Law and AI. Faculty of Law and Political Science, June 4, 2024.
- Marco Winckler
 - "Information Visualization and Analytical Provenance: the missing gap" at the international seminars in relations to Human Computer Interaction, Software Engineering and Accessibility, at KTH Visualization studio, Stockholm, Sweden, June 14th 2024.
 - "Journée dédiée aux technologies immersives et à l’apprentissage", organized by France Immersive Learning, in cooperation with EducAzur and CCI Nice Cote d’Azur, at the Campus Sud des Métiers, Nice, April 3rd 2024. (round table)
 - Inria-Brasil Workshop on AI and Applications. “Contributing to the research in AI and the (Semantic) Web”. Hybrid event physically held at LNCC, Petropolis, Brazil on April 11, 2024.

11.1.5 Scientific expertise

- Elena Cabrio: Member of the Academic Board of University Côte d'Azur, and of the board of Académie 1 IDEX UniCA JEDI.
- Catherine Faron: Member of the HCERES committee in charge of evaluating Paris Cité 's training offer; Member of the HCERES committee in charge of evaluating Paris 8's training offer; Member of the HCERES committee in charge of evaluating the LIASD laboratory; Expert for the European commission, Horizon MSSCA and CL4 programmes.
- Fabien Gandon: European Science Foundation (ESF): Evaluation of an FWO research project.
- Serena Villata: Member of the Bureau of CEP of Inria SAM, Member and vice-chair of the CE23 project evaluation committee of ANR on "Artificial Intelligence", member of the HCERES committee evaluating LTCI Laboratory (IPP), member of the scientific committee of DGFiP.
- Marco Winckler: Member of the XRC2 Center for Extended Reality Université Côte d'Azur. Steering Committee Chair for the IFIP TC13 Conference INTERACT.

11.1.6 Research administration

- Fabien Gandon : Leader of the Wimmics team ; co-president of scientific and pedagogical council of the Data Science Technical Institute (**DSTI**), W3C Advisory Committee Representative (AC Rep) for Inria.
- Serena Villata: Deputy scientific director of 3IA Côte d'Azur.
- Marco Winckler : Leader of the SPARKS team of the CNRS laboratory I3S (UMR 7271).

11.1.7 Promoting Open Science practices

Too often, the methods described in research papers are not reproducible because the code and/or data are simply not provided. As a result, it is hardly possible to verify the results and build upon these works. The Open Science movement is meant to fix this by fostering the unhindered spreading of the results, methods and products of scientific research. It is based on the open access to publications, data and source codes.

To make the team members aware of these issues, in 2024 we gave a 2-hour presentation on the principles of Open Science, the goals of experiment reproducibility, with a focus on practical approaches meant to make codes and data findable (using metadata), accessible (public repositories and long-time preservation), referenceable (point to a specific version) and citable (give credit, attribution), as well as good practices to cite others' codes and data.

Franck Michel. Open Science, reproducible research, and the citation of articles, code and data alike. 2024. [Slides](#).

11.2 Teaching - Supervision - Juries

11.2.1 Teaching

Participants: Michel Buffa, Elena Cabrio, Olivier Corby, Catherine Faron, Fabien Gandon, Aline Menin, Amaya Nogales Gómez, Andrea Tettamanzi, Serena Villata, Marco Winckler, Molka Dhouib, Benjamin Molinet, Célian Ringwald, Pierre Monnin, Anaïs Ollagnier.

- Michel Buffa:
 - Licence 3, Master 1, Master 2 Méthodes Informatiques Appliquées à la Gestion des Entreprises (MIAGE) : Web Technologies, Web Components, etc. 192h.

- DS4H Masters 3D games programming on Web, JavaScript Introduction: 40h.
- Olivier Corby:
 - Licence 3 IOTA UniCA 25 hours Semantic Web
 - Licence 3 IA DS4H UniCA 25 hours Semantic Web
- Catherine Faron :
 - Master 2/5A SI PNS: Web of Data, 32 h
 - Master 2/5A SI PNS: Semantic Web 32h
 - Master 2/5A SI PNS: Ingénierie des connaissances 15h
 - Master DSAI UniCA: Web of Data, 30h
 - Master 1/4A SI PNS and Master2 IMAFA/5A MAM PNS: Web languages, 28h
 - Licence 3/3A SI PNS and Master 1/4A MAM PNS: Relational Databases, 60h
 - Master **Data ScienceTech Institute (DSTI)**: Data pipeline, 50h.
- Fabien Gandon :
 - Master: Integrating Semantic Web technologies in Data Science developments, 72 h, M2, **Data ScienceTech Institute (DSTI)**, France.
 - Tutorial Inria Academy, “CORESE and Semantic Web Standards”, 6h, twice : June and December
- Aline Menin :
 - Master 2, Data Visualization, 12h éq. TD (CM/TD), UniCA, MBDS DS4H, France.
 - Polytech 5ème année, UniCA, 13.5h (CM/TP), Data visualization.
 - BUT 2, IUT Nice Côte d’Azur, 160h éq. TD (CM/TD), “Développement efficace”, “Qualité de développement”, and “Développement des Applications avec IHM”.
- Molinet Benjamin:
 - Fabron BUT-1 - Intro NLP - 42h TD.
 - EMSI Casablanca, Master IA2 - Natural Language Processing, 20h CM/TD
 - Master I Computer Science, Text Processing in AI, 2 hours.
- Serena Villata:
 - Master II Droit de la Création et du Numérique - Sorbonne University: Approche de l’Elaboration et du Fonctionnement des Logiciels, 15 hours (CM), 20 students.
 - Master 2 MIAGE IA - University Côte d’Azur: I.A. et Langage : Traitement automatique du langage naturel, 28 hours (CM+TP), 30 students.
 - DUT IA et santé. Natural Language Processing, 4 hours.
- Elena Cabrio:
 - Master 1 Computer Science, Text Processing in AI, 30 hours.
 - Master 2 MIAGE IA: I.A. et Langage : Traitement automatique du langage naturel, 8 hours.
 - Master 1 EUR CREATES, Parcours Linguistique, traitements informatiques du texte et processus cognitifs. Introduction to Computational Linguistics, 30 hours.
 - Master 1 EUR CREATES, Parcours Linguistique, traitements informatiques du texte et processus cognitifs. Textual Data Analysis 30 hours.
 - DUT IA et santé. Natural Language Processing, 2 hours.

- Licence 2 IUT, Introduction to AI, 30 hours.
- Licence 1 IUT, Introduction to Database and SQL, 100 hours.
- Andrea Tettamanzi
 - Licence: Introduction à l'Intelligence Artificielle, 45 h ETD, L2, UniCA, France.
 - Master: Logic for AI, 30 h ETD, M1, UniCA, France.
 - Master: Web, 30 h ETD, M1, UniCA, France.
 - Master: Algorithmes Évolutionnaires, 24.5 h ETD, M2, UniCA, France.
 - Master: Modélisation del l'Incertitude, 24.5 h ETD, M2, UniCA, France.
- Marco Winckler
 - Licence 3: Event-driven programming, 45 h ETD, UniCA, Polytech Nice, France.
 - Master 1: Methods and tools for technical and scientific writing, Master DSAI, 15 h ETD, UniCA, DS4H, France.
 - Master 2: Introduction to Scientific Research, 15 h ETD, UniCA, Polytech Nice, France.
 - Master 2: Information Visualisation, 34 h ETD, UniCA, Polytech Nice, France.2024
 - Master 2: Data Visualization, 15 h ETD, UniCA, MBDS DS4H, France.
 - Master 2: Design of Interactive Systems, 34 ETD, UniCA, Polytech Nice, France.
 - Master 2: Evaluation of Interactive Systems, 34 ETD, UniCA, Polytech Nice, France.
 - Master 2: Multimodal Interaction Techniques, 15 ETD, UniCA, Polytech Nice, France.
 - Master 2: Coordination of the TER (Travaux de Fin d'Etude), UniCA, Polytech Nice, France.
 - Master 2: Coordination of the track on Human-Computer Interaction at the Informatics Department, UniCA, Polytech Nice, France.
- Molka Dhouib
 - Licence 3/3A SI PNS: Relational Databases, 32,5h (TD).
 - Master 1/4A SI PNS: Web languages, 10h (TD)
 - Licence 3/LPI: Introduction to Web of Data and Semantic Web, 16h (TD)
- Célian Ringwald:
 - Licence 3/3A SI PNS: Base de données relationnelles, 63,5 h (TD) .
- Pierre Monnin:
 - Master 2/Applied Foreign Languages: Artificial Intelligence, professional applications. 4h CM, 4h TD
 - Master 1/Applied Foreign Languages: Introduction to Artificial Intelligence. 9h CM, 9h TD
 - Master 1/Adult Education: Introduction to Artificial Intelligence. 24h CM
 - Bachelor 1/Preparatory classes for Higher Education: Introduction to Artificial Intelligence. 6h CM
- Anaïs Ollagnier:
 - Master 1/2/Phd EUR LEXSOCIETE. Introduction to AI Applied to Law and Implications for Administration and Public Service. 10h CM
 - Master 1/2/Phd EUR LEXSOCIETE. Introduction to AI Applied to Law. 12h CM, 8h TD
 - Master 1/2/Phd EUR ELMI. Artificial Intelligence and societal tranformation. 12h CM
- Arnaud Barbe:

- 1ere Année BUT : Introduction aux bases de données relationnelles et SQL, 40h TD.
- Franck Michel:
 - Master 2/5A SI PNS: Web of Data, 16 h

E-learning

- Mooc: Fabien Gandon, Olivier Corby & Catherine Faron, Web of Data and Semantic Web (FR), 7 weeks, [FUN](#), Inria, France Université Numérique, self-paced course 41002, Education for Adults, 17496 learners registered at the time of this report and 855 certificates/badges, [MOOC page](#).
- Mooc: Fabien Gandon, Olivier Corby & Catherine Faron, Introduction to a Web of Linked Data (EN), 4 weeks, [FUN](#), Inria, France Université Numérique, self-paced course 41013, Education for Adults, 5952 learners registered at the time of this report, [MOOC page](#).
- Mooc: Fabien Gandon, Olivier Corby & Catherine Faron, Web of Data (EN), 4 weeks, [Coursera](#), self-paced course Education for Adults, 5134 learners registered at the time of this report, [MOOC page](#).
- Mooc: Michel Buffa, HTML5 coding essentials and best practices, 6 weeks, [edX MIT/Harvard](#), self-paced course Education for Adults, more than 500k learners at the time of this report (2015-2024), [MOOC page](#).
- Mooc: Michel Buffa, HTML5 Apps and Games, 5 weeks, [edX MIT/Harvard](#), self-paced course Education for Adults, more than 150k learners at the time of this report (2015-2024), [MOOC page](#).
- Mooc: Michel Buffa, JavaScript Introduction, 5 weeks, [edX MIT/Harvard](#), self-paced course Education for Adults, more than 250k learners at the time of this report (2015-2024), [MOOC page](#).

11.2.2 Supervision

PhDs

- Defended PhD: Ali Ballout, *Active Learning for Axiom Discovery*, Supervised by Andrea Tettamanzi, UniCA.
- Defended PhD: Lucie Cadorel, *Localisation sur le territoire et prise en compte de l'incertitude lors de l'extraction des caractéristiques de biens immobiliers à partir d'annonces*, Supervised by Andrea Tettamanzi, UniCA.
- Defended PhD: Rony Dupuy Charles, *Combinaison d'approches symboliques et connexionnistes d'apprentissage automatique pour les nouvelles méthodes de recherche et développement en agro-végétale-environnement*, Supervised by Andrea Tettamanzi, UniCA.
- Defended PhD: Rémi Felin, *Découverte évolutive d'axiomes à partir de graphes de connaissances*, UniCA, Co-supervised by Andrea Tettamanzi, Catherine Faron.
- Defended PhD: Pierpaolo Goffredo, *Fallacious Argumentation in Political Debates*, UniCA 3IA, Co-supervised by Elena Cabrio and Serena Villata.
- PhD in progress: Guillaume Méroué, *Decentralized coordination of intelligent processes to improve the quality of a collection of knowledge graphs*, UniCA 3IA, Co-supervised by Pierre Monnin and Fabien Gandon.
- Defended PhD: Benjamin Molinet, *Explanatory argument generation for healthcare applications*, UniCA 3IA, Co-supervised by Elena Cabrio and Serena Villata.
- PhD in progress: Benjamin Ocampo, *Subtle and Implicit Hate Speech Detection*, UniCA 3IA, Co-supervised by Elena Cabrio and Serena Villata.

- PhD in progress: Clément Quere, *Immersive Visualization Techniques for spatial-temporal data*. Co-supervised by Aline Menin, Hui-Yin Wu, and Marco Winckler.
- PhD in progress: Célian Ringwald, *Learning RDF pattern extractors for a language from dual bases Wikipedia/LOD*, UniCA 3IA, Co-supervised by Fabien Gandon, Catherine Faron, Franck Michel, Hanna Abi-Akl.
- PhD in progress: Nicolas Robert, *Knowledge graph embedding models: symbolic knowledge injection and discovery*, UniCA, Co-supervised by Catherine Faron and Pierre Monnin.
- Defended PhD: Florent Robert, *Analyzing and Understanding Embodied Interactions in Extended Reality Systems*. Co-supervised by Hui-Yin Wu, Lucile Sassatelli, and Marco Winckler.
- Defended PhD: Maroua Tikat, *Visualisation multimédia interactive pour l'exploration d'une base de métadonnées multidimensionnelle de musiques populaires*. Co-supervised by Michel Buffa, Marco Winckler.
- PhD in progress: Xiaou Wang, *Counter-argumentation generation to fight online disinformation*, UniCA, Co-supervised by Elena Cabrio and Serena Villata.
- PhD in progress: Greta Damo, *Argument-based counter narratives generation to fight online hate speech*, UniCA, Co-supervised by Elena Cabrio and Serena Villata.
- PhD in progress: Ekaterina Sviridova, *Unveiling Implicit Inferences for Advanced Argument Extraction*, UniCA, Co-supervised by Elena Cabrio and Serena Villata.
- PhD in progress: Deborah Dore, *Mining arguments over time*, UniCA, Co-supervised by Elena Cabrio and Serena Villata.
- PhD in progress: Cyprien Michel-Deletie, *Automatic Evaluation of Consistency and Reliability of Information Sources : Toward Deeper Understanding of Discourse with Language Models and Logical Analysis*, UniCA, Co-supervised by Elena Cabrio and Serena Villata.

Internships and Apprenticeships

- Apprentice (M2 NumRes) Hajar Bakarou, “Enhancing Fusion Embedding Techniques for Hate Speech Detection in Conversational Data”, Supervised by Anaïs Ollagnier.
- Apprentice (M2 NumRes) Mohamed Sinane El Messoussi, “Infusing Pragmatic Knowledge for Online Hate Detection”, Supervised by Anaïs Ollagnier.
- Apprentice (M1 MIAGE) Erwan Hain, “AI-boosted Deliberation”, Supervised by Elena Cabrio.
- Internship (Polytech, M2) Guillaume Mérroué. “Analogical Reasoning for Link Prediction in Knowledge Graphs”. Co-supervised by Pierre Monnin and Fabien Gandon.

Master Projects (TER/PER Projet d'Etude et de Recherches Polytech)

- Master 2 PER, Polytech: Guillaume Mérroué. Supervised by Pierre Monnin
- Master 2 Informatique PER + Gratification DS4H: Dheeraj Parkash. “Exploring Signed Network Embedding for Hate Speech Detection on Conversational Data” supervised by Anaïs Ollagnier
- Master 2 Informatique PER + Gratification DS4H: Ahmed Muhammad. “Enhancing Argument Mining through Contextual Information Synthesis” supervised by Anaïs Ollagnier
- Master 2 PER (M2 IA2) Hajar Bakarou, “Enhancing Fusion Embedding Techniques for Hate Speech Detection in Conversational Data”, Supervised by Anaïs Ollagnier.
- Master 2 PER (M2 IA2) Mohamed Sinane El Messoussi, “Exploring Signed Network Embedding for Hate Speech Detection on Conversational Data”, Supervised by Anaïs Ollagnier.

- Master 2 PER (M2 Polytech) Quentin Scordo, “Computing and visualizing innovative bibliometric indicators for the articles of a scientific archive”, Supervised by Franck Michel.
- Master 2 PER (M2 Polytech) Quentin BOURDEAU, Ambre CORREIA et Théo JEANNE, “The Use of Polymorphic Glyphs to Support the Exploration of Spatiotemporal Data”. Co-supervised by Aline Menin and Marco Winckler.
- Master 2 PER (M2 Polytech) Christophe RUIZ and Enzo DAVAL, “Interactive Visualization and Allocation of Office Spaces”. Co-supervised by Aline Menin and Marco Winckler.
- Master 1 PER (M2 DS4H) Steven ESSAM EDWAR AZIZ, “Design and development of a tool for the preservation, archiving, and exhibition of extended reality projects”. Supervised by Aline Menin.

11.2.3 Juries

- Catherine Faron
 - reviewer of the HDR of Sébastien Harispe, Ecole des Mines d’Alès
 - reviewer of the HDR of Lylia Gouaich-Abrouk, Université de Bourgogne
 - reviewer of the PhD thesis of Julian Bruyat, Université de Lyon
 - reviewer of the PhD thesis of Yousouf Taghzouti, Ecole des Mines de Saint Etienne
 - Member of the PhD thesis jury of Mohamed Lechiakh, University Mohammed VI Polytechnic, Morocco
 - President of the PhD jury of Maroua Tikat, Université Côte d’Azur
 - President of the PhD jury of Dupuis Rony Charles, Université Côte d’Azur
 - President of the PhD jury of Benjamin Molinet, Université Côte d’Azur
 - Member of the jury for the PhD prize at the French conference on data management BDA 2024
 - Member of the Nice committee of Inria center at Université Côte d’Azur
 - Member of the I3S laboratory board
- Fabien Gandon
 - President of the jury of the PhD of Ali Ballout on *Active Learning for Axiom Discovery*
 - President of the jury of the PhD of Pierpaolo Goffredo, *Fallacious Argumentation in Political Debates*
 - President of the jury of the PhD of Rémi Felin on *Découverte évolutive d’axiomes à partir de graphes de connaissances*
- Serena Villata
 - Reviewer of the PhD thesis entitled “Knowledge-Enhanced Natural Language Processing” by Giacomo Frisoni, University of Bologna, April 2024.
 - Reviewer of the PhD thesis entitled “Evaluating and Improving the Reasoning Abilities of Language Models” by Chadi Helwé, LIX, June 2024.
 - Reviewer of the PhD thesis entitled “Advancing Fairness in Natural Language Processing: From Traditional Methods to Explainability” by Fanny Jourdan, University of Toulouse, June 2024.
- Elena Cabrio
 - Reviewer of the PhD thesis entitled "Computational approaches to language Change" by Pierluigi Cassotti, University of Bari (Italy), February 2024.
 - Reviewer of the PhD thesis entitled "Boolean Algebra and Random Field Theory applied to Graph Neural Network and Natural Language Processing - Application to Argument Mining in a low resource setting", by Samuel Guilluy, Rennes, March 2024.

- Reviewer of the PhD thesis entitled "Summarizing User-generated Discourse" by Shahbaz Syed, University of Leipzig (Germany), March 2024.
 - Reviewer of the PhD thesis entitled "Multimodal Extraction of Proofs and Theorems from the Scientific Literature" by Shrey Mishra, ENS-PSL, June 2024.
 - Reviewer of the PhD thesis entitled "Détection des valeurs humaines dans les commentaires des consommateurs sur les produits parfumés", by Boyu NIU, INALCO Paris, September 2024.
 - Reviewer of the PhD thesis entitled "Generic Framework for the Multidimensional Processing and Analysis of Social Media Content - A Proxemic Approach" by Maxime Masson, University of Pau, September 2024.
 - Reviewer of the PhD thesis entitled "Viewpoints Detection in Political Speeches" by Tu My Doan Norwegian University of Science and Technology, September 2024.
 - Reviewer of the PhD thesis entitled "Improving Trust in Fact-Checking Systems with Synthetic Training Data and Explanations" by Jean-Flavien Bussotti Pitollet, EURECOM, November 2024.
- Marco Winckler
 - Reviewer of the PhD thesis entitled "Acceptabilité de l'intelligence artificielle en contexte professionnel: Facteurs d'influence et méthodologies d'évaluation" by Alexandre AGOSSAH. Nantes Université, France. October 9th, 2024.
 - Member of the jury of the PhD thesis entitled "A Framework for Digital Inclusion" by Dena HUSSEIN AL-OMRAN, KTH, Stockholm, Sweden, June 14th, 2024.
 - Member of the jury of the Individual Monitoring Committee (CSI second year) for the PhD thesis entitled "Exploration et analyse interactive de grands graphes de données" by Théo BOUGANIM, Université Paris-Saclay, France, October 15th, 2024.
 - Member of the Recrutement Committee for Associated Professor (MCF 490) at Université Polytech Paris-Saclay, France. Member (rapporteur) of the committee.
 - Member of the Recrutement Committee for Associated Professor Associated Professor (MCF 233) at Université Côte d'Azur, France. Member (rapporteur) of the committee.
 - Member of the Recrutement Committee for Associated Professor Associated Professor (MCF 4880) at Université Paul Sabatier, France. President* of the committee.
 - Member of the Recrutement Committee for Associated Professor Associated Professor (MCF 4925) at IUT Castre/Université Paul Sabatier.
 - Aline Menin
 - Invited Member of Maroua Tikat's PhD defense jury. The thesis is entitled *Visualisation multimédia interactive pour l'exploration d'une base de métadonnées multidimensionnelle de musiques populaires*.

11.3 Popularization

11.3.1 Productions (articles, videos, podcasts, serious games, ...)

- Interview of Fabien Gandon in "Des Initiatives Pour Redonner Confiance en L'IA", Science & Cerveau, September - November 2024, P:80-89
- Interview of Fabien Gandon "Intelligence artificielle : peut-on avoir confiance en elle ?", Sciences et avenir, June 2024

11.3.2 Participation in Live events

- Elena Cabrio: participation at the Science Festival 2024 in Nice, October 2024.
- Serena Villata: participation at the Science Festival 2024 in Antibes, October 2024. Participation to the round table at the Printemps des technologies 2024, March 2024, Centre de Congres Ville Saint Raphael, L'Intelligence Artificielle dans nos vies - dernieres applications.
- Marco Winckler: participation at the "Journée d'accélération de l'ICCARE-Lab : Demain: quel audiovisuel?", November 29th, 2024, Quartier de la Création, Nantes, France. (round table)

12 Scientific production

12.1 Major publications

- [1] D. Allemang, J. Hendler and F. Gandon. *Semantic Web for the Working Ontologist*. 3. ACM, June 2020. DOI: [10.1145/3382097](https://doi.org/10.1145/3382097). URL: <https://hal.inria.fr/hal-02939606> (cit. on p. 7).
- [2] A. Ballout. 'Active learning for axiom discovery'. Université Côte d'Azur, 24th June 2024. URL: <https://theses.hal.science/tel-04680907>.
- [3] A. Ben othmane. 'CARS - A multi-agent framework to support the decision making in uncertain spatio-temporal real-world applications'. Université Côte d'Azur, 12th Oct. 2017. URL: <https://hal.archives-ouvertes.fr/tel-01680984>.
- [4] F. Berthelon. 'Emotion modelization and detection from expressive and contextual data'. Université Nice Sophia Antipolis, 16th Dec. 2013. URL: <https://tel.archives-ouvertes.fr/tel-00917416>.
- [5] K. R. Bouzidi. 'Semantic web models to support the creation of technical regulatory documents in building industry'. Université Nice Sophia Antipolis, 11th Sept. 2013. URL: <https://tel.archives-ouvertes.fr/tel-00876366>.
- [6] E. Cabrio. 'Artificial Intelligence to Extract, Analyze and Generate Knowledge and Arguments from Texts to Support Informed Interaction and Decision Making'. Habilitation à diriger des recherches. Université Côte d'Azur, Oct. 2020. URL: <https://hal.inria.fr/tel-03084380>.
- [7] L. Cadorel. 'Qualifying and quantifying uncertainty of geolocation information extracted from french real estated ads'. Université Côte d'Azur, 24th Jan. 2024. URL: <https://theses.hal.science/tel-04524701>.
- [8] M. Chaves, E. Cabrio and S. Villata. 'FALCON: A multi-label graph-based dataset for fallacy classification in the COVID-19 infodemic'. In: SAC '25 - ACM/SIGAPP Symposium on Applied Computing. Catania, Italy, 2025. DOI: [10.1145/3672608.3707913](https://doi.org/10.1145/3672608.3707913). URL: <https://hal.science/hal-04834405>.
- [9] L. Costabello. 'Context-aware access control and presentation of linked data'. Université Nice Sophia Antipolis, 29th Nov. 2013. URL: <https://tel.archives-ouvertes.fr/tel-00934617>.
- [10] P. F. Diallo. 'Sociocultural and temporal aspects in ontologies dedicated to virtual communities'. COMUE Université Côte d'Azur (2015 - 2019); Université de Saint-Louis (Sénégal), 16th Sept. 2016. URL: <https://tel.archives-ouvertes.fr/tel-01402394>.
- [11] A. E. A. Djebri. 'Uncertainty Management for Linked Data Reliability on the Semantic Web'. Université Côte D'Azur, 24th Feb. 2022. URL: <https://hal.archives-ouvertes.fr/tel-03679118>.
- [12] A. Ettore. 'Towards an interpretable model of learners in a learning environment based on knowledge graphs'. Université Côte d'Azur, 28th Nov. 2022. URL: <https://theses.hal.science/tel-04060324>.
- [13] R. Felin. 'Evolutionary knowledge discovery from RDF data graphs'. Université Côte d'Azur, 22nd Nov. 2024. URL: <https://theses.hal.science/tel-04874737> (cit. on p. 39).

- [14] M. Fell. ‘Natural language processing for music information retrieval : deep analysis of lyrics structure and content’. Theses. Université Côte d’Azur, May 2020. URL: <https://tel.archives-ouvertes.fr/tel-02587910>.
- [15] F. Gandon. ‘Distributed Artificial Intelligence And Knowledge Management: Ontologies And Multi-Agent Systems For A Corporate Semantic Web’. Université Nice Sophia Antipolis, 7th Nov. 2002. URL: <https://tel.archives-ouvertes.fr/tel-00378201>.
- [16] R. Gazzotti. ‘Knowledge graphs based extension of patients’ files to predict hospitalization’. Theses. Université Côte d’Azur, Apr. 2020. URL: <https://hal.archives-ouvertes.fr/tel-02959420>.
- [17] P. Goffredo, E. Cabrio, S. Villata, S. Haddadan and J. Torres Sanchez. ‘DISPUTool 2.0: A Modular Architecture for Multi-Layer Argumentative Analysis of Political Debates’. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. AAAI Conference on Artificial Intelligence. Vol. 37. 13. Washington, DC, United States, 26th June 2023, pp. 16431–16433. DOI: [10.1609/aaai.v37i13.27069](https://doi.org/10.1609/aaai.v37i13.27069). URL: <https://hal.science/hal-04349796> (cit. on p. 32).
- [18] P. Goffredo, M. Espinoza, S. Villata and E. Cabrio. ‘Argument-based Detection and Classification of Fallacies in Political Debates’. In: *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. SINGAPORE, Singapore: Association for Computational Linguistics, Dec. 2023, pp. 11101–11112. DOI: [10.18653/v1/2023.emnlp-main.684](https://doi.org/10.18653/v1/2023.emnlp-main.684). URL: <https://hal.science/hal-04260221> (cit. on p. 33).
- [19] P. Goffredo, S. Haddadan, V. Vorakitphan, E. Cabrio and S. Villata. ‘Fallacious Argument Classification in Political Debates’. In: *Thirty-First International Joint Conference on Artificial Intelligence IJCAI-22*. Vienna, Austria: International Joint Conferences on Artificial Intelligence Organization, 23rd July 2022, pp. 4143–4149. DOI: [10.24963/ijcai.2022/575](https://doi.org/10.24963/ijcai.2022/575). URL: <https://hal.science/hal-03873412>.
- [20] N. Halliwell. ‘Evaluating and improving explanation quality of graph neural network link prediction on knowledge graphs’. Université Côte d’Azur, 23rd Nov. 2022. URL: <https://theses.hal.science/tel-03907696>.
- [21] R. Hasan. ‘Predicting query performance and explaining results to assist Linked Data consumption’. Université Nice Sophia Antipolis, 4th Nov. 2014. URL: <https://tel.archives-ouvertes.fr/tel-01127124>.
- [22] M. Lefrançois. ‘Meaning-Text Theory lexical semantic knowledge representation : conceptualization, representation, and operationalization of lexicographic definitions’. Université Nice Sophia Antipolis, 24th June 2014. URL: <https://tel.archives-ouvertes.fr/tel-01071945>.
- [23] A. Macina. ‘SPARQL distributed query processing over linked data’. COMUE Université Côte d’Azur (2015 - 2019), 17th Dec. 2018. URL: <https://tel.archives-ouvertes.fr/tel-02340700>.
- [24] N. Marie. ‘Linked data based exploratory search’. Université Nice Sophia Antipolis, 12th Dec. 2014. URL: <https://tel.archives-ouvertes.fr/tel-01130622>.
- [25] T. Mayer. ‘Argument Mining on Clinical Trials’. Université Côte d’Azur, 16th Dec. 2020. URL: <https://tel.archives-ouvertes.fr/tel-03209489>.
- [26] Z. Meng. ‘Temporal and semantic analysis of richly typed social networks from user-generated content sites on the web’. Université Côte d’Azur, 7th Nov. 2016. URL: <https://hal.inria.fr/tel-01402612>.
- [27] F. Michel. ‘Integrating heterogeneous data sources in the Web of data’. Université Côte d’Azur, 3rd Mar. 2017. URL: <https://tel.archives-ouvertes.fr/tel-01508602>.
- [28] F. Michel, F. Gandon, V. Ah-Kane, A. Bobasheva, E. Cabrio, O. Corby, R. Gazzotti, A. Giboin, S. Marro, T. Mayer, M. Simon, S. Villata and M. Winckler. ‘Covid-on-the-Web: Knowledge Graph and Services to Advance COVID-19 Research’. In: *ISWC 2020 - 19th International Semantic Web Conference*. Athens / Virtual, Greece, 1st Nov. 2020. DOI: [10.1007/978-3-030-62466-8_19](https://doi.org/10.1007/978-3-030-62466-8_19). URL: <https://hal.archives-ouvertes.fr/hal-02939363>.
- [29] T. H. Nguyen. ‘Mining the semantic Web for OWL axioms’. Université Côte d’Azur, 2nd July 2021. URL: <https://tel.archives-ouvertes.fr/tel-03406784>.

- [30] C. Pasquier, C. Da Costa Pereira and A. G. B. Tettamanzi. ‘Extending a Fuzzy Polarity Propagation Method for Multi-Domain Sentiment Analysis with Word Embedding and POS Tagging’. In: *Frontiers in Artificial Intelligence and Applications*. ECAI 2020 - 24th European Conference on Artificial Intelligence. Vol. 325. Santiago de Compostela, Spain: IOS Press, 29th Aug. 2020, pp. 2140–2147. DOI: [10.3233/FAIA200338](https://doi.org/10.3233/FAIA200338). URL: <https://hal.inria.fr/hal-02936130>.
- [31] T. A. Pham. ‘OntoApp : a declarative approach for software reuse and simulation in early stage of software development life cycle’. Université Côte d’Azur, 22nd Sept. 2017. URL: <https://tel.archives-ouvertes.fr/tel-01680766>.
- [32] O. Seye. ‘Sharing and reusing rules for the Web of data’. Université Nice Sophia Antipolis; Université Gaston Berger de Saint Louis, 15th Dec. 2014. URL: <https://hal.inria.fr/tel-01096306>.
- [33] M. Tounsi Dhouib. ‘Knowledge engineering in the sourcing domain for the recommendation of providers’. Université Côte d’Azur, 26th Mar. 2021. URL: <https://tel.archives-ouvertes.fr/tel-03336353>.
- [34] M. Toure. ‘Local peer-to-peer mobile access to linked data in resource-constrained networks’. Université Côte d’Azur; Université de Saint-Louis (Sénégal), 4th Oct. 2021. URL: <https://tel.archives-ouvertes.fr/tel-03514271>.
- [35] D. M. Tran. ‘Discovering multi-relational association rules from ontological knowledge bases to enrich ontologies’. Université Côte d’Azur; Université de Danang (Vietnam), 23rd July 2018. URL: <https://tel.archives-ouvertes.fr/tel-01926812>.

12.2 Publications of the year

International journals

- [36] S. Bistarelli, V. David, F. Santini and C. Taticchi. ‘Temporal Duration-Based Probabilistic Argumentation Frameworks’. In: *Journal of Logic and Computation* (31st July 2024). DOI: [10.1093/logcom/exae039](https://doi.org/10.1093/logcom/exae039). URL: <https://inria.hal.science/hal-04851343> (cit. on p. 44).
- [37] J. Bruyat, P.-A. Champin, L. Médini and F. Laforest. ‘PRSC: from PG to RDF and back, using schemas’. In: *Semantic Web – Interoperability, Usability, Applicability* (2024). URL: <https://hal.science/hal-04772450>.
- [38] C. A. Martins, C. F. Dorneles and M. A. Winckler. ‘A Comprehensive Review of User Interaction for Recommendation Systems’. In: *iSys - Brazilian Journal of Information Systems* (Dec. 2024). DOI: [10.5753/isys.2024.4064](https://doi.org/10.5753/isys.2024.4064). URL: <https://inria.hal.science/hal-04877488>.
- [39] A. Menin, C. Quere, J. Wagner, S. Chardonnel, P.-A. Davoine, W. Stuerzlinger, C. M. Dal Sasso Freitas, L. Nedel and M. Winckler. ‘What about thematic information? An analysis of the multidimensional visualization of individual mobility’. In: *Visual Informatics* (Feb. 2025). DOI: [10.1016/j.visinf.2025.02.002](https://doi.org/10.1016/j.visinf.2025.02.002). URL: <https://hal.science/hal-04944676>. In press.
- [40] B. Molinet, S. Marro, E. Cabrio and S. Villata. ‘Explanatory argumentation in natural language for correct and incorrect medical diagnoses’. In: *Journal of Biomedical Semantics* 15.8 (2024), p. 8. DOI: [10.1186/s13326-024-00306-1](https://doi.org/10.1186/s13326-024-00306-1). URL: <https://hal.science/hal-04602527> (cit. on p. 38).
- [41] N. Yacoubi Ayadi, S. Bernard, R. Bossy, M. Courtin, B. G. Happi Happi, P. Larmande, F. Michel, C. Nedellec, C. Roussey and C. Faron. ‘A Unified approach to publish semantic annotations of agricultural documents as knowledge graphs’. In: *Smart Agricultural Technology* 8 (31st Jan. 2024), p. 43. DOI: [10.1016/j.atech.2024.100484](https://doi.org/10.1016/j.atech.2024.100484). URL: <https://hal.science/hal-04495022> (cit. on pp. 27, 37).

National journals

- [42] M. Buffa and A. Vidal-Mazuy. ‘Une station de travail audio-numérique open-source pour la plate-forme Web’. In: *Revue Francophone d’Informatique et Musique* 10.10 (20th Aug. 2024). DOI: [10.56698/rfim.746](https://doi.org/10.56698/rfim.746). URL: <https://inria.hal.science/hal-04695266>.

- [43] P. Monnin, C.-H. Noursadine, L. Jarnac, L. Zuckerman and M. Couceiro. ‘KGPrune : une application Web pour extraire des sous-graphes d’intérêt de Wikidata par élagage analogique’. In: *1024 : Bulletin de la Société Informatique de France* 24 (Nov. 2024), pp. 177–188. DOI: [10.48556/SIF.1024.24.177](https://doi.org/10.48556/SIF.1024.24.177). URL: <https://inria.hal.science/hal-04854745> (cit. on p. 43).

International peer-reviewed conferences

- [44] H. Abi Akl. ‘DSTI at LLMs4OL 2024 Task A: Intrinsic versus extrinsic knowledge for type classification: Applications on WordNet and GeoNames datasets’. In: *TIB Open Access Publishing proceedings*. 1st LLMs4OL Challenge @ ISWC 2024. Maryland / USA, United States, 1st Oct. 2024. URL: <https://hal.science/hal-04678365>.
- [45] H. A. Akl. ‘NeSy is alive and well: A LLM-driven symbolic approach for better code comment data generation and classification’. In: *The First International Workshop on Generative Neuro-Symbolic AI (GeNeSy) @ ESWC 2024*. ESWC 2024 - Extended Semantic Web Conference. Hersonissos, Greece, 30th May 2024. URL: <https://hal.science/hal-04481420>.
- [46] H. A. Akl. ‘Project SHADOW: Symbolic Higher-order Associative Deductive reasoning On Wikidata using LM probing’. In: *NATL 2024 - 10th International Conference on Natural Language Computing*. Melbourne, Australia, 10th Nov. 2024. URL: <https://hal.science/hal-04706047>.
- [47] H. A. Akl, D. Mariko, Y.-A. Pilatte, S. Durfort, N. Yahiaoui and A. Gupta. ‘Knowledge Graphs can play together: Addressing knowledge graph alignment from ontologies in the biomedical domain’. In: *KDIR 2024 - 16th International Conference on Knowledge Discovery and Information Retrieval*. Porto, Portugal, 17th Nov. 2024. URL: <https://hal.science/hal-04708946>.
- [48] A. Ballout, C. da Costa Pereira and A. G. B. Tettamanzi. ‘Scalable Prediction of Atomic Candidate OWL Class Axioms Using a Vector-Space Dimension Reduced Approach’. In: *dblp computer science bibliography*. ICAART 2024 - 16th International Conference on Agents and Artificial Intelligence. Vol. 3. 16th ICAART 2024, vol 3. Rome, Italy: SCITEPRESS - Science and Technology Publications, 2024, pp. 347–357. DOI: [10.5220/0012384200003636](https://doi.org/10.5220/0012384200003636). URL: <https://inria.hal.science/hal-04567587> (cit. on p. 39).
- [49] S. Bistarelli, V. David, P. Monnin, F. Santini and C. Taticchi. ‘Fast Computing of Dung Semantics in Acyclic Probabilistic Argumentation Frameworks’. In: *The 39th Annual AAAI Conference on Artificial Intelligence (AAAI 2025)*. Thirty-Ninth AAAI Conference on Artificial Intelligence, AAAI 2025. Philadelphia, United States, 25th Feb. 2025. URL: <https://inria.hal.science/hal-04876258> (cit. on p. 44).
- [50] H. Bonaldi, G. Damo, N. B. Ocampo, E. Cabrio, S. Villata and M. Guerini. ‘Is Safer Better? The Impact of Guardrails on the Argumentative Strength of LLMs in Hate Speech Countering’. In: *EMNLP 2024 - 2024 Conference on Empirical Methods in Natural Language Processing*. Miami (Florida), United States: arXiv, 2024. DOI: [10.48550/arXiv.2410.03466](https://doi.org/10.48550/arXiv.2410.03466). URL: <https://hal.science/hal-04734455> (cit. on p. 32).
- [51] M. Buffa and S. Demont. ‘Can you DAW it Online?’ In: *IS2 2024 - IEEE International Symposium on the Internet of Sounds 2024 / 1st IEEE International Workshop on the Musical Metaverse (IEEE IWMM)*. Erlangen, Germany, 29th Sept. 2024. URL: <https://inria.hal.science/hal-04695237> (cit. on p. 29).
- [52] M. Buffa, D. Girard, S. Demont, Q. Escobar and A. Hofr. ‘Faust Plugins in (Sometimes Unexpected) Web-Based Hosts’. In: *International Faust Conference 2024*. Turin, Italy, 21st Nov. 2024. URL: <https://inria.hal.science/hal-04733381> (cit. on p. 29).
- [53] M. Buffa, A. Hofr and D. Girard. ‘Using Web Audio Modules for Immersive Audio Collaboration in the Musical Metaverse’. In: *IS2 2024 - IEEE International Symposium on the Internet of Sounds 2024*. Erlangen, Germany, 30th Sept. 2024. URL: <https://inria.hal.science/hal-04695228> (cit. on p. 29).
- [54] M. Buffa, S. Ren, T. Burns, A. Vidal-Mazuy and S. Letz. ‘Evolution of the Web Audio Modules Ecosystem’. In: *Zenodo, WAC community*. WAC 2024 - Web Audio Conference 2024. Lafayette, Indiana, United States: Zenodo, 15th Mar. 2024. DOI: [10.5281/zenodo.10825647](https://doi.org/10.5281/zenodo.10825647). URL: <https://inria.hal.science/hal-04507622> (cit. on p. 29).

- [55] C. Cardellino, T. A. Collias, B. Molinet, E. Hain, W. Sun, R. Agerri, S. Villata and E. Cabrio. ‘ANTI-DOTE: Argumentation-Driven explainable artificial intelligence for digital medicine’. In: ECAI 2024 Demos Proceedings - 27th European Conference on Artificial Intelligence. Santiago de Compostela, Spain, 19th Oct. 2024. URL: <https://hal.science/hal-04673974> (cit. on p. 38).
- [56] D. Charles, P. Pultrini and A. G. B. Tettamanzi. ‘Outlier Detection in MET Data Using Subspace Outlier Detection Method’. In: *dblp computer science bibliography*. ICAART 2024 - 16th International Conference on Agents and Artificial Intelligence. 16th ICAART 2024: Rome, Italy - Volume 3. Rome, Italy: SCITEPRESS - Science and Technology Publications, 24th Feb. 2024, pp. 243–250. DOI: [10.5220/0012318000003636](https://doi.org/10.5220/0012318000003636). URL: <https://inria.hal.science/hal-04567556> (cit. on p. 40).
- [57] M. Chaves, E. Cabrio and S. Villata. ‘FALCON: A multi-label graph-based dataset for fallacy classification in the COVID-19 infodemic’. In: SAC ’25 - ACM/SIGAPP Symposium on Applied Computing. Catania, Italy, 2025. DOI: [10.1145/3672608.3707913](https://doi.org/10.1145/3672608.3707913). URL: <https://hal.science/hal-04834405> (cit. on pp. 28, 33).
- [58] G. Damo, N. B. Ocampo, E. Cabrio and S. Villata. ‘PEACE: Providing Explanations and Analysis for Combating Hate Expressions’. In: ECAI 2024 - 27th European Conference on Artificial Intelligence. Santiago de Compostela, Spain, 19th Oct. 2024. URL: <https://hal.science/hal-04684950> (cit. on pp. 13, 31).
- [59] G. Damo, N. B. Ocampo, E. Cabrio and S. Villata. ‘Unveiling the Hate: Generating Faithful and Plausible Explanations for Implicit and Subtle Hate Speech Detection’. In: NLDB 2024 - 29th International Conference on Natural Language & Information Systems. Torino, Italy, 25th June 2024. URL: <https://hal.science/hal-04658110> (cit. on p. 31).
- [60] R. Felin, P. Monnin, C. Faron and A. G. B. Tettamanzi. ‘An Algorithm Based on Grammatical Evolution for Discovering SHACL Constraints’. In: EuroGP 2024 - 27th European Conference on Genetic Programming. Vol. 14631. Genetic Programming – 27th European Conference, EuroGP 2024. Aberystwyth, United Kingdom: Springer Nature Switzerland, 2024, pp. 176–191. DOI: [10.1007/978-3-031-56957-9_11](https://doi.org/10.1007/978-3-031-56957-9_11). URL: <https://inria.hal.science/hal-04446252> (cit. on p. 39).
- [61] I. García-Ferrero, R. Agerri, A. Atutxa, E. Cabrio, I. de la Iglesia, A. Lavelli, B. Magnini, B. Molinet, J. Ramirez-Romero, G. Rigau, J. M. Villa-Gonzalez, S. Villata and A. Zaninello. ‘Medical mT5: An Open-Source Multilingual Text-to-Text LLM for The Medical Domain’. In: LREC-COLING 2024 - 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation. TORINO, Italy, 20th May 2024. URL: <https://hal.science/hal-04603270> (cit. on p. 38).
- [62] M. Gille Levenson, C. Ringwald, M. Flesch, J. Isasi, S. Papastamkou, R. Quiroga and D. Valentine. ‘Connecter les chapitres linguistiques de Programming Historian ? : Premières ébauches d’une table conceptuelle multilingue constituée semi-automatiquement’. In: HAL. Humanistica 2024. Sciences de l’information. Meknès, Morocco, 3rd May 2024. URL: <https://hal.science/hal-04557817>.
- [63] F.-Z. H. Hannou, V. Charpenay, M. Lefrançois, C. Roussey, A. Zimmermann and F. Gandon. ‘La méthodologie ACIMOV pour l’intégration agile et continue des modules ontologiques’. In: IC 2024 - 35èmes Journées francophones d’Ingénierie des Connaissances - Plateforme d’intelligence artificielle (PFIA 2024). La Rochelle, France, 1st July 2024, pp. 127–128. URL: <https://hal-emse.ccsd.cnrs.fr/emse-04636109> (cit. on p. 37).
- [64] N. Hubert, P. Monnin, A. Brun and D. Monticolo. ‘Treat Different Negatives Differently: Enriching Loss Functions with Domain and Range Constraints for Link Prediction’. In: *Lecture Notes in Computer Science*. ESWC 2024 - 21st International Conference on Semantic Web. Semantic Web - 21st International Conference, ESWC 2024. Hersonissos, Greece, 26th May 2024. URL: <https://inria.hal.science/hal-04491230> (cit. on pp. 13, 42).

- [65] N. Hubert, P. Monnin, M. D'aquin, D. Monticolo and A. Brun. 'PyGraft: Configurable Generation of Synthetic Schemas and Knowledge Graphs at Your Fingertips'. In: *Lecture notes in computer science*. ESWC 2024 - 21st International Conference on Semantic Web. Semantic Web - 21st International Conference, ESWC 2024. Hersonissos, Greece, 26th May 2024. DOI: [10.5281/zenodo.10243209](https://doi.org/10.5281/zenodo.10243209). URL: <https://inria.hal.science/hal-04491258> (cit. on pp. 13, 42).
- [66] P. Maillot, J. Andersen, S. Cazalens, C. Faron, F. Gandon, P. Lamarre and F. Michel. 'An Open Platform for Quality Measures in a Linked Data Index'. In: WWW '24: The ACM Web Conference 2024. Singapore, Singapore: ACM, 13th May 2024, pp. 1087–1090. DOI: [10.1145/3589335.3651443](https://doi.org/10.1145/3589335.3651443). URL: <https://inria.hal.science/hal-04575211> (cit. on p. 41).
- [67] C. A. Martins, C. Dorneles, A. Barišić, T. R. Silva and M. Winckler. 'CM-DIR: A Method to Support the Specification of the User's Dynamic Behavior in Recommender Systems'. In: *Lecture notes in computer science*. HCSE 2024 - International Conference on Human-Centred Software Engineering. Vol. LNCS-14793. Human-Centered Software Engineering 10th IFIP WG 13.2 International Working Conference, HCSE 2024, Reykjavik, Iceland, July 8–10, 2024, Proceedings 14793. Reykjavik, Iceland: Springer, 2024, pp. 26–46. DOI: [10.1007/978-3-031-64576-1_2](https://doi.org/10.1007/978-3-031-64576-1_2). URL: <https://inria.hal.science/hal-04829247>.
- [68] F. Michel and F. Gandon. 'Pay Attention: a Call to Regulate the Attention Market and Prevent Algorithmic Emotional Governance'. In: *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. AIES 2024 - 7th AAAI/ACM Conference on AI, Ethics, and Society. Vol. 7 (2024): Proceedings of the Seventh AAAI/ACM Conference on AI, Ethics, and Society (AIES-24) 7. San Jose (CA), United States: AAAI Press, Washington, DC, USA, 17th Oct. 2024. URL: <https://hal.science/hal-04758276> (cit. on p. 35).
- [69] H. Nasser, D. El Zein, C. da Costa Pereira, C. Escazut and A. G. B. Tettamanzi. 'RULKKG: Estimating User's Knowledge Gain in Search-as-Learning Using Knowledge Graphs'. In: *ACM Digital Library*. CHIIR 2024 : 2024 ACM SIGIR Conference on Human Information Interaction and Retrieval. CHIIR '24: Proceedings of the 2024 Conference on Human Information Interaction and Retrieval. Sheffield - UK, United Kingdom: ACM, 10th Mar. 2024, pp. 364–369. DOI: [10.1145/3627508.3638331](https://doi.org/10.1145/3627508.3638331). URL: <https://inria.hal.science/hal-04567545> (cit. on p. 40).
- [70] A. Ollagnier. 'CyberAgressionAdo-v2: Leveraging Pragmatic-Level Information to Decipher Online Hate in French Multiparty Chats'. In: The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation. Torino, Italy, 20th May 2024. URL: <https://hal.science/hal-04514689> (cit. on p. 26).
- [71] M. A. N. Pour, A. Algergawy, E. Blomqvist, P. Buche, J. Chen, P. G. Cotovio, A. Coulet, J. Cufi, H. Dong, D. Faria, L. Ferraz, S. Hertling, Y. He, I. Horrocks, L. Ibanescu, D. S. Jain, E. Jiménez-Ruiz, N. Karam, F. Kraus, P. Lambrix, H. Li, Y. Li, P. Monnin, H. Paulheim, C. Pesquita, A. Sharma, P. Shvaiko, M. Silva, G. Sousa, C. Trojahn, J. Vataščinová, B. Yaman, O. Zamazal and L. Zhou. 'Results of the Ontology Alignment Evaluation Initiative 2024'. In: *CEUR workshop proceedings*. OM 2024 - 19th International Workshop on Ontology Matching (collocated with ISWC-2024). Vol. CEUR-3897. Proceedings of the 19th International Workshop on Ontology Matching co-located with the 23rd International Semantic Web Conference (ISWC 2024). Baltimore, United States, 2024, pp. 1–34. URL: <https://inria.hal.science/hal-04892635>.
- [72] C. Quere, A. Menin, R. Julien, H.-Y. Wu and M. Winckler. 'HandyNotes: using the hands to create semantic representations of contextually aware real-world objects'. In: *IEEE Xplore*. IEEE VR 2024 - The 31st IEEE Conference on Virtual Reality and 3D User Interfaces. Orlando, Florida, United States, 16th Mar. 2024. URL: <https://inria.hal.science/hal-04425616> (cit. on p. 28).
- [73] S. Ren, S. Letz, M. Buffa, L. Pottier and Y. Yu. 'Design and Run Real-time Spectral Processing on the Web with Faust'. In: *Zenodo*. WAC 2024 - Web Audio Conference 2024. Lafayette, Indiana, United States: Zenodo, 15th Mar. 2024. DOI: [10.5281/zenodo.10825715](https://doi.org/10.5281/zenodo.10825715). URL: <https://inria.hal.science/hal-04507625> (cit. on p. 29).

- [74] C. Ringwald. ‘Learning Pattern-Based Extractors from Natural Language and Knowledge Graphs: Applying Large Language Models to Wikipedia and Linked Open Data’. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. AAAI-24 - 38th AAAI Conference on Artificial Intelligence. Vol. 38. Vol. 38 No. 21: IAAI-24, EAAI-24, AAAI-24 Student Abstracts, Undergraduate Consortium and Demonstrations 21. Vancouver, Canada, 24th Mar. 2024, pp. 23411–23412. DOI: [10.1609/aaai.v38i21.30406](https://doi.org/10.1609/aaai.v38i21.30406). URL: <https://hal.science/hal-04526050>.
- [75] C. Ringwald, F. Gandon, C. Faron, F. Michel and H. Akl. ‘Well-Written Knowledge Graphs: Most Effective RDF Syntaxes for Triple Linearization in End-to-End Extraction of Relations from Texts (Student Abstract): Well-Written Knowledge Graphs’. In: *Proceedings of the 38th AAAI Conference on Artificial Intelligence*. AAAI 2024 - 38th Annual AAAI Conference on Artificial Intelligence. Vol. 38. 21. Vancouver, Canada, 27th Feb. 2024, pp. 23631–23632. DOI: [10.1609/aaai.v38i21.30502](https://doi.org/10.1609/aaai.v38i21.30502). URL: <https://hal.science/hal-04384724>.
- [76] C. Ringwald, F. Gandon, C. Faron, F. Michel and H. Akl. ‘12 shades of RDF: Impact of Syntaxes on Data Extraction with Language Models’. In: *ESWC 2024 Extended Semantic Web Conference*. Hersonissos, Greece, 26th May 2024. URL: <https://hal.science/hal-04581124> (cit. on pp. 13, 41).
- [77] F. A. S. Robert, H.-Y. Wu, L. Sassatelli and M. Winckler. ‘Task-based methodology to characterise immersive user experience with multivariate data’. In: *IEEE VR 2024 - 31st IEEE conference on virtual reality and 3D user interfaces*. Orlando (FL), United States, 16th Mar. 2024. URL: <https://hal.science/hal-04446066>.
- [78] S. Shen, M. Treleani, A. Menin and M. Winckler. ‘Les Échos visuels: Exploration des Phénomènes de Répétition dans les Archives audiovisuelles’. In: *HAL. Humanistica 2024. Sciences de l’information*. Meknès, Morocco, 3rd May 2024. URL: <https://hal.science/hal-04563421>.
- [79] E. Sviridova, A. Yeginbergen, A. Estarrona, E. Cabrio, S. Villata and R. Agerri. ‘CasiMedicos-Arg: A Medical Question Answering Dataset Annotated with Explanatory Argumentative Structures’. In: *ACL Anthology*. EMNLP 2024 - Conference on Empirical Methods in Natural Language Processing. Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing. Miami, United States: Association for Computational Linguistics, Nov. 2024, pp. 18463–18475. DOI: [10.18653/v1/2024.emnlp-main.1026](https://doi.org/10.18653/v1/2024.emnlp-main.1026). URL: <https://hal.science/hal-04823453> (cit. on p. 35).
- [80] V. Vieira, E. Oliveira, A. Menin, E. Clua, M. Winckler and D. Trevisan. ‘Understanding affordances in XR interactions through a design space’. In: *ACM Digital Library*. IHC 2024 - XXIII Brazilian Symposium on Human Factors in Computing Systems. Vol. 2024-10-11. IHC ’24: Proceedings of the XXIII Brazilian Symposium on Human Factors in Computing Systems. Brasília DF Brazil, France: ACM, 7th Oct. 2024, pp. 1–14. DOI: [10.1145/3702038.3702112](https://doi.org/10.1145/3702038.3702112). URL: <https://inria.hal.science/hal-04875209> (cit. on p. 28).
- [81] H. Wachsmuth, G. Lapesa, E. Cabrio, A. Lauscher, J. Park, E. M. Vecchi, S. Villata and T. Ziegenbein. ‘Argument Quality Assessment in the Age of Instruction-Following Large Language Models’. In: *ACL Anthology*. LREC/COLING 2024 - Joint International Conference on Computational Linguistics, Language Resources and Evaluation, LREC/COLING 2024. Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). Torino, Italy, 20th May 2024. URL: <https://hal.science/hal-04787971> (cit. on p. 32).
- [82] X. Wang, E. Cabrio and S. Villata. ‘Argument-structured Justification Generation for Explainable Fact-checking’. In: *IEEE Xplore*. WI-IAT 2024 - 23rd IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology. Bangkok, Thailand, 9th Dec. 2024. URL: <https://inria.hal.science/hal-04862965> (cit. on p. 34).

National peer-reviewed Conferences

- [83] F. Amardeilh, S. Aubin, S. Bernard, S. Bravo, R. Bossy, C. Faron, F. Michel and C. Roussey. ‘Aligner les descriptions des plantes ayant des points de vue distincts’. In: *IC2024 : 35es Journées francophones d’Ingénierie des Connaissances*. 35es Journées francophones d’Ingénierie des Connaissances (IC

- 2024) @ Plate-Forme Intelligence Artificielle (PFIA 2024). Vol. IC. Plate-Forme Intelligence Artificielle (PFIA) 2024. La Rochelle, France: AFIA-Association Française pour l'Intelligence Artificielle, 1st July 2024, pp. 121–122. DOI: [10.57745/LVRFWJ](https://doi.org/10.57745/LVRFWJ). URL: <https://hal.science/hal-04638307> (cit. on p. 37).
- [84] R. Felin, P. Monnin, C. Faron and A. G. B. Tettamanzi. 'Extraction probabiliste de formes SHACL à l'aide d'algorithmes évolutionnaires'. In: *Revue des Nouvelles Technologies de l'Information*. EGC 2024 - Extraction et Gestion de la Connaissance. Vol. RNTI-E-40. Dijon, France, 2024. URL: <https://inria.hal.science/hal-04411349> (cit. on p. 39).
- [85] N. Hubert, P. Monnin, A. Brun and D. Monticolo. 'Enrichissement de fonctions de perte avec contraintes de domaine et co-domaine pour la prédiction de liens dans les graphes de connaissance'. In: *Actes de la 27e Conférence Nationale en Intelligence Artificielle*. 27e Conférence Nationale en Intelligence Artificielle. La Rochelle, France, 1st July 2024. URL: <https://inria.hal.science/hal-04566996> (cit. on p. 42).
- [86] N. Hubert, P. Monnin, M. D'aquin, D. Monticolo and A. Brun. 'PyGraft: un outil Python pour la génération de schémas et graphes de connaissance synthétiques'. In: 35es Journées francophones d'Ingénierie des Connaissances (IC 2024) @ Plate-Forme Intelligence Artificielle (PFIA 2024). IC2024 : 35es Journées francophones d'Ingénierie des Connaissances. La Rochelle, France, 3rd July 2024. URL: <https://inria.hal.science/hal-04589855> (cit. on p. 42).
- [87] S. Ouelhadj, P.-A. Champin and J. Gaillard. 'sETL: Outils ETL pour la construction de graphes de connaissances en exploitant la sémantique implicite des schémas de données'. In: *Journées francophones d'Ingénierie des Connaissances (IC 2024) @ Plate-Forme Intelligence Artificielle (PFIA 2024)*. IC20 2024 - 35èmes journées francophones d'Ingénierie des Connaissances. Événement affilié à PFIA 2024 - Plate-Forme Intelligence Artificielle. IC2024 : 35es Journées francophones d'Ingénierie des Connaissances. La Rochelle, France, 1st July 2024. URL: <https://inria.hal.science/hal-04635968>.
- [88] M. Tounsi Dhouib, E. Kostyrykina and C. Faron. 'Automatic definition of the level of textual difficulty of documents'. In: *Revue des Nouvelles Technologies de l'Information*. 24ème conférence francophone sur l'Extraction et la Gestion des Connaissances. Dijon, France, 2024. URL: <https://hal.science/hal-04353060>.

Conferences without proceedings

- [89] V. David, R. Fournier-S'Niehotta and N. Travers. 'MAP Inference Reasoning on TMLN with Neo4j'. In: BDA'24. Orléans, France, 21st Oct. 2024. URL: <https://hal.science/hal-04753923> (cit. on p. 44).
- [90] J. Ferrari-Giovanangeli, M. Treleani and A. Menin. 'ILIADÉ – Héritage immersif: Les enjeux des standards et de l'institutionnalisation manquée'. In: Journées thématiques Arts et Humanités 2024. Nice, France, 21st Mar. 2024. URL: <https://hal.science/hal-04637223>.

Edition (books, proceedings, special issue of a journal)

- [91] M. Couceiro, E. Marquer, P. Monnin and P.-A. Murena, eds. *Preface to the special issue on analogies: from mathematical foundations to applications and interactions with ML and AI* (28th Dec. 2024). DOI: [10.1007/s10472-024-09961-7](https://doi.org/10.1007/s10472-024-09961-7). URL: <https://inria.hal.science/hal-04858719>.
- [92] C. Faron and S. Loudcher, eds. *Special issue on Advances in Knowledge Discovery and Management, Best Papers of EGC 2023*. 154.Special issue (Dec. 2024). URL: <https://hal.science/hal-04804382>.
- [93] *Engineering Interactive Computer Systems. EICS 2023 International Workshops and Doctoral Consortium : Swansea, UK, June 26-27, 2023, Selected Papers*. EICS 2023 - 15th ACM SIGCHI Symposium on Engineering Interactive Computing Systems. Vol. LNCS-14517. Lecture Notes in Computer Science. Springer Nature Switzerland, 2024, pp. I–XIV, 1–224. DOI: [10.1007/978-3-031-59235-5](https://doi.org/10.1007/978-3-031-59235-5). URL: <https://hal.science/hal-04678911>.

Doctoral dissertations and habilitation theses

- [94] A. Ballout. 'Active learning for axiom discovery'. Université Côte d'Azur, 24th June 2024. URL: <https://theses.hal.science/tel-04680907> (cit. on p. 39).
- [95] L. Cadorel. 'Qualifying and quantifying uncertainty of geolocation information extracted from french real estated ads'. Université Côte d'Azur, 24th Jan. 2024. URL: <https://theses.hal.science/tel-04524701> (cit. on p. 39).
- [96] B. Molinet. 'Argument-based natural language explanation generation and assessment in health-care'. Université Côte d'Azur, 18th Dec. 2024. URL: <https://theses.hal.science/tel-04952437>.
- [97] F. Robert. 'Analyzing and understanding embodied interactions in virtual reality systems'. Université Côte d'Azur, 9th Dec. 2024. URL: <https://theses.hal.science/tel-04913490>.
- [98] M. Tikat. 'Contributing to the interactive annotation of multi-dimensional knowledge graphs : a case study of popular music data'. Université Côte d'Azur, 18th Sept. 2024. URL: <https://theses.hal.science/tel-04932453>.

Reports & preprints

- [99] J. Ben-Naim, V. David and A. Hunter. *An Axiomatic Study of the Evaluation of Enthymeme Decoding in Weighted Structured Argumentation*. 13th Nov. 2024. URL: <https://inria.hal.science/hal-04851405> (cit. on p. 43).
- [100] J. Ben-Naim, V. David and A. Hunter. *Understanding Enthymemes in Argument Maps: Bridging Argument Mining and Logic-based Argumentation*. 16th Aug. 2024. URL: <https://inria.hal.science/hal-04851392> (cit. on p. 43).

Other scientific publications

- [101] M. Buffa and S. Demont. 'DAW it again! An Audio Workstation for the Web'. In: IS2 2024 - IEEE International Symposium on the Internet of Sounds 2024. Erlangen, Germany, 29th Sept. 2024. URL: <https://inria.hal.science/hal-04695244> (cit. on p. 29).
- [102] M. Buffa, A. Hofr and D. Girard. 'WAM Jam Party: Using Web Audio Modules in the Musical Metaverse'. In: IS2 2024 - IEEE International Symposium on the Internet of Sounds 2024 / 1st IEEE International Workshop on the Musical Metaverse (IEEE IWMM). Erlangen, Germany, 29th Sept. 2024. URL: <https://inria.hal.science/hal-04695252> (cit. on p. 29).
- [103] C. Cardellino, S. Villata and E. Cabrio. 'From Political Debates to Deliberative Democracy: A Roadmap to Assess Semi-Supervised Argument Mining with DISPUTool'. In: DELITE 2024T - the First Workshop on Language-driven Deliberation Technology. Torino, Italy, 20th May 2024. URL: <https://hal.science/hal-04567307> (cit. on p. 34).
- [104] R. Felin, P. Monnin, C. Faron and A. G. B. Tettamanzi. 'RDFminer: an Interactive Tool for the Evolutionary Discovery of SHACL Shapes'. In: ESWC 2024 - 21st International Conference on Semantic Web. Hersonissos (Crete), Greece, 26th May 2024. URL: <https://inria.hal.science/hal-04566981> (cit. on p. 39).
- [105] F. Gandon. *Knowledge Graphs as the Foundation for Interoperable Intelligent Systems: Keynote Fabien Gandon KGSWC 2024*. Université Paris Cité, France, 11th Dec. 2024. URL: <https://inria.hal.science/hal-04838584> (cit. on p. 55).
- [106] F. Metereau, F. Michel, P. LARMANDE, G. Sempere and C. Faron. 'Integration of variation data through SPARQL Micro-Services'. In: Semantic Web for Health Care and Life Science (SWA4HCLS). Leiden (DE), Netherlands, 26th Feb. 2024. URL: <https://ird.hal.science/ird-04812730>.

- [107] P. Monnin, C.-H. Noursradine, L. Jarnac, L. Zuckerman and M. Couceiro. ‘KGPrune: a Web Application to Extract Subgraphs of Interest from Wikidata with Analogical Pruning’. In: ECAI 2024 - 27th European Conference on Artificial Intelligence. ECAI 2024 - 27th European Conference on Artificial Intelligence, October 19-24, 2024, Santiago de Compostela, Spain. Santiago de Compostela, Spain: IOS Press, 16th Oct. 2024. DOI: [10.3233/FAIA241038](https://doi.org/10.3233/FAIA241038). URL: <https://inria.hal.science/hal-04678284> (cit. on p. 43).
- [108] C. Ringwald, F. Gandon, C. Faron, F. Michel and H. A. Akl. ‘Learning Pattern-Based Extractors from Natural Language and Knowledge Graphs Applying Large Language Models to Wikipedia & the Linked Open Data (POSTER)’. In: AAAI 2024 - 38th Annual AAAI Conference on Artificial Intelligence. Vancouver, France, 20th Feb. 2024. URL: <https://hal.science/hal-04526139> (cit. on p. 41).
- [109] C. Ringwald, F. Gandon, C. Faron, F. Michel and H. A. Akl. ‘Well-written Knowledge Graphs Most Effective RDF Syntaxes for Triple Linearization in End-to-End Extraction of Relations from Text (Student Abstract)’. In: AAAI 24 - 38th Annual AAAI Conference on Artificial Intelligence. Vancouver, Canada, 20th Feb. 2024. URL: <https://hal.science/hal-04526132> (cit. on p. 41).

Software

- [110] [SW] A. Bobasheva, F. Michel, A. Tchechmedjiev and A. Toulet, *ISSA Pipeline* version 2.1.0, 27th Nov. 2024. Inria & Université Cote d’Azur, CNRS, I3S, Sophia Antipolis, France; Cirad; Euromov Digital Health in Motion, IMT Mines Alès. LIC: Apache License 2.0. DOI: [10.5281/zenodo.1423046](https://doi.org/10.5281/zenodo.1423046), HAL: [hal-04807540](https://hal.science/hal-04807540), URL: <https://hal.science/hal-04807540>, VCS: <https://github.com/issa-project/issa-pipeline>, SWHID: [sw:1:rev:0ca5a450c3ba350eb2bb413a7b1fdc7aa2a293b](https://sw.hal.science/sw/1:rev:0ca5a450c3ba350eb2bb413a7b1fdc7aa2a293b) (cit. on p. 40).
- [111] [SW] N. Robert, F. Gandon and M. Lefrançois, *olivaw* version 0.0.7, 6th Dec. 2024. LIC: <https://spdx.org/licenses/LGPL-2.1>. DOI: [10.5281/zenodo.14288084](https://doi.org/10.5281/zenodo.14288084), HAL: [hal-04823080](https://hal.science/hal-04823080), URL: <https://hal.science/hal-04823080>, VCS: <https://github.com/Wimmics/olivaw.git>, SWHID: [sw:1:dir:92e5f4f833d5c1082fe1e3c2d0d931663906258f](https://sw.hal.science/sw/1:dir:92e5f4f833d5c1082fe1e3c2d0d931663906258f) (cit. on p. 37).

12.3 Cited publications

- [112] A. Barbe, M. Tounsi Dhouib, C. Faron, M. Corneli and A. Zucker. ‘Construction d’un graphe de connaissance à partir des annotations manuelles de textes de zoologie antique’. In: *IC2023 : 34es Journées francophones d’Ingénierie des Connaissances*. IC2023 : 34es Journées francophones d’Ingénierie des Connaissances. Starsbourg, France, July 2023. URL: <https://inria.hal.science/hal-04156996> (cit. on p. 36).
- [113] [SW Rel.] R. Cérés, O. Corby, E. Demairy and F. Gandon, *Corese* version 4.4.1, July 2023. Inria ; CNRS ; Université Côte d’Azur. LIC: CECILL-C. DOI: [10.5281/zenodo.8181618](https://doi.org/10.5281/zenodo.8181618), HAL: [hal-04170333v1](https://hal.science/hal-04170333v1), URL: <https://hal.science/hal-04170333>, VCS: <https://github.com/Wimmics/corese>, SWHID: [sw:1:dir:a6072722ea23608d1bdd95d1377f5c5159558cba](https://sw.hal.science/sw/1:dir:a6072722ea23608d1bdd95d1377f5c5159558cba) (cit. on p. 37).
- [114] R. Felin, C. Faron and A. G. B. Tettamanzi. ‘A Framework to Include and Exploit Probabilistic Information in SHACL Validation Reports’. In: *Lecture Notes in Computer Science*. Vol. LNCS-13870. The Semantic Web 20th International Conference, ESWC 2023, Hersonissos, Crete, Greece, May 28–June 1, 2023, Proceedings. Hersonissos, Greece: Springer Nature Switzerland, May 2023, pp. 91–104. DOI: [10.1007/978-3-031-33455-9_6](https://doi.org/10.1007/978-3-031-33455-9_6). URL: <https://inria.hal.science/hal-04031744> (cit. on p. 39).
- [115] F. Gandon. ‘The three ‘W’ of the World Wide Web call for the three ‘M’ of a Massively Multidisciplinary Methodology’. In: *10th International Conference, WEBIST 2014*. Ed. by V. Monfort and K.-H. Krempels. Vol. 226. Web Information Systems and Technologies. Barcelona, Spain: Springer International Publishing, Apr. 2014. DOI: [10.1007/978-3-319-27030-2](https://doi.org/10.1007/978-3-319-27030-2). URL: <https://hal.inria.fr/hal-01223236> (cit. on p. 7).

- [116] F. Gandon, M. Buffa, E. Cabrio, O. Corby, C. Faron Zucker, A. Giboin, N. Le Thanh, I. Mirbel, P. Sander, A. G. B. Tettamanzi and S. Villata. ‘Challenges in Bridging Social Semantics and Formal Semantics on the Web’. In: *5h International Conference, ICEIS 2013*. Ed. by S. Hammoudi, J. Cordeiro, L. Maciaszek and J. Filipe. Vol. 190. Angers, France: Springer, July 2013, pp. 3–15. URL: <https://hal.inria.fr/hal-01059273> (cit. on p. 7).
- [117] P. Goffredo, M. Chaves, S. Villata and E. Cabrio. ‘Argument-based Detection and Classification of Fallacies in Political Debates’. In: *ACL Anthology*. Vol. 2023.findings-emnlp.684. Findings of the Association for Computational Linguistics: EMNLP 2023. Singapore (SG), Singapore: Association for Computational Linguistics, Dec. 2023, pp. 11101–11112. DOI: [10.18653/v1/2023.emnlp-main.684](https://doi.org/10.18653/v1/2023.emnlp-main.684). URL: <https://hal.science/hal-04260221> (cit. on p. 26).
- [118] P. Maillot, O. Corby, C. Faron, F. Gandon and F. Michel. ‘IndeGx: A Model and a Framework for Indexing RDF Knowledge Graphs with SPARQL-based Test Suites’. In: *Journal of Web Semantics* (Jan. 2023). DOI: [10.1016/j.websem.2023.100775](https://doi.org/10.1016/j.websem.2023.100775). URL: <https://hal.science/hal-03946680> (cit. on pp. 30, 41).
- [119] F. Michel, O. Gargominy, S. Tercerie and C. Faron Zucker. ‘A Model to Represent Nomenclatural and Taxonomic Information as Linked Data. Application to the French Taxonomic Register, TAXREF’. In: *ISWC 2017 Workshop on Semantics for Biodiversity (S4Biodiv 2017)*. Vol. CEUR Vol. 1933. Vienna, Austria, Oct. 2017, pp. 1–12. URL: <https://hal.science/hal-01617708> (cit. on p. 26).
- [120] I. Pajón Leyra, A. Zucker and C. Faron Zucker. ‘Thezoo: un thesaurus de zoologie ancienne et médiévale pour l’annotation de sources de données hétérogènes’. In: *Archivum Latinitatis Medii Aevi* 73 (2015), pp. 321–342. URL: <https://hal.science/hal-01335652> (cit. on p. 36).
- [121] S. Pamerlon, A.-S. Archambeau, F. Pando de la Hoz, G. Olivier, F. Michel, S. Tercerie, E. Rodinson, N. Maurel, G. Martínez-Sagarra, A. Kerner, R. Vignes Lebbe, B. Schatz and P. Dupont. ‘Knowledge Base on Species Life Traits : A Spanish/French Plinian Core implementation use case’. In: *Biodiversity Information Science and Standards* 7 (Aug. 2023), e111784. DOI: [10.3897/biss.7.111784](https://doi.org/10.3897/biss.7.111784). URL: <https://hal.science/hal-04198414> (cit. on p. 37).
- [122] M. Tounsi Dhouib, Q. Merilleau, C. Guerrero, M. Corneli, C. Faron and A. Zucker. ‘Automatic Semantic Classification of Ancient Zoological Texts’. In: *Journal of cultural Heritage*. Nice, France, Nov. 2023. URL: <https://hal.science/hal-04210669> (cit. on p. 36).
- [123] E. Tysinger, M. Pagni, O. Kirchhoffer, F. Mehl, F. Gandon, J.-L. Wolfender and L.-F. Nothias. ‘An Artificial Intelligence Agent for Navigating Knowledge Graph Experimental Metabolomics Data’. In: *Frontiers in Metabolomics - Book of abstract -*. Frontiers in Metabolomics. Swiss Metabolomics Society Zurich and ETH Zurich. Zurich, Switzerland, Sept. 2023. URL: <https://hal.science/hal-04381448> (cit. on p. 42).