RESEARCH CENTRE

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

IN PARTNERSHIP WITH:

CNRS, Université Côte d'Azur

2022
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

Creation of the Project-Team: 2013 July 01

Keywords

Computer sciences and digital sciences

A1.2.9. – Social Networks

A1.3.1. – Web

A1.3.4. – Peer to peer

A2.1. – Programming Languages

A2.1.1. – Semantics of programming languages

A3.1.1. – Modeling, representation

A3.1.2. – Data management, quering and storage

A3.1.3. – Distributed data

A3.1.4. – Uncertain data

A3.1.5. – Control access, privacy

A3.1.6. – Query optimization

A3.1.7. – Open data

A3.1.9. – Database

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

A3.3.1. – On-line analytical processing

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, HDR]

- Olivier Corby [INRIA, Researcher]

- Damien Graux [INRIA, ISFP, until Dec 2022]

- Yusuke Higuchi [AIST JAPAN, Researcher, from Jul 2022]

- Serena Villata [CNRS, Senior Researcher, HDR]

Faculty Members

- Marco Alba Winckler [UNIV COTE D'AZUR, Professor, HDR]

- Michel Buffa [UNIV COTE D'AZUR, Professor, HDR]

- Elena Cabrio [UNIV COTE D'AZUR, Professor, HDR]

- Pierre-Antoine Champin [UNIV LYON, Associate Professor, Delegation, HDR]

- Catherine Faron [UNIV COTE D'AZUR, Professor, HDR]

- Nhan Le Thanh [UNIV COTE D'AZUR, Professor, Emeritus, HDR]

- Aline Menin [UNIV COTE D'AZUR, Associate Professor, from Sep 2022]

- Amaya Nogales Gomez [UNIV COTE D'AZUR, Associate Professor, 3IA short-term position]

- Peter Sander [UNIV COTE D'AZUR, Professor, until Aug 2022, HDR]

- Andrea Tettamanzi [UNIV COTE D'AZUR, Professor, HDR]

Post-Doctoral Fellows

- Pierre Maillot [INRIA]

- Anaïs Ollagnier [UNIV COTE D'AZUR]

- Nadia Yacoubi Ayadi [CNRS]

PhD Students

- Ali Ballout [UNIV COTE D'AZUR]

- Lucie Cadorel [KINAXIA, CIFRE]

- Rony Dupuy Charles [DORIANE, CIFRE]

- Antonia Ettorre [UNIV COTE D'AZUR]

- Remi Felin [UNIV COTE D'AZUR]

- Pierpaolo Goffredo [CNRS]

- Mina Ilhan [UNIV COTE D'AZUR]

- Santiago Marro [CNRS]

- Benjamin Molinet [UNIV COTE D'AZUR]

- Nicolas Ocampo [UNIV COTE D'AZUR]

- Clement Quere [UNIV COTE D'AZUR, from Oct 2022]

- Shihong Ren [UNIV JEAN MONNET, Saint-Étienne]

- Florent Robert [Université Côte d'Azur, Co-supervision with Hui-Yin Wu (BIOVISION INRIA Team)]

- Maroua Tikat [UNIV COTE D'AZUR]

- Xiaoou Wang [CNRS, from Jul 2022]

Technical Staff

- Anna Bobasheva [INRIA, Engineer]

- Remi Ceres [INRIA, Engineer]

- Molka Dhouib [INRIA, Engineer]

- Maxime Lecoq [INRIA, Engineer, from Oct 2022, Plan relance, hosted by Startinblox]

- Christopher Leturc [INRIA, Engineer]

- Franck Michel [CNRS, Engineer]

- Iliana Petrova [INRIA, Engineer]

- Celian Ringwald [INRIA, Engineer]

- Ekaterina Sviridova [UNIV COTE D'AZUR, Engineer, from Sep 2022]

Interns and Apprentices

- Poulomi Guha [UNIV COTE D'AZUR]

- Martina Rossini [CNRS, from Sep 2022]

- Ekaterina Sviridova [UNIV COTE D'AZUR, until Jul 2022]

- Antoine Vidal Mazuy [INRIA, from Sep 2022, Apprentice]

Administrative Assistants

- Christine Foggia [INRIA]

- Lionel Tavanti [UNIV COTE D'AZUR]

External Collaborators

- Andrei Ciortea [UNIV ST GALLEN, Assistant Professor]

- Alain Giboin [Retired, Emeritus]

- Freddy Lecue [J.P. MORGAN, AI Research Director]

- Oscar Rodriguez Rocha [TEACH ON MARS]

- 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[101].

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[102], 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 knowledge graphs: 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 meaningul 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: 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 as typed graphs the models identified in the previous objectives and to populate thems 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 semantic graphs on the Web: 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 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 Representation and Extraction 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 organisations 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, Mona-LIA

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 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, .

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 and SPARQL 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 her current mental and physical state, the emotion she just expressed or 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 her personal information and to link it to public or collective one, maintaining 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 contributed 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 argurment 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). I

5   Highlights of the year

5.1   General news

- Catherine Faron and Elena Cabrio were promoted Full Professor (Université Côte d'Azur).

- Serena Villata was promoted Research Director (CNRS).

- Marco Winckler become head of the SPARKS team of I3S/CNRS.

- Fabien Gandon became co-president of the scientific board of the Data ScienceTech Institute.

- Pierre-Antoine Champin joined the team to become a W3C Fellow and support the "Data activity" in the consortium.

5.2   Defenses

We had three PhD defenses in the team:

- Antonia Ettorre on an "Interpretable model of learners in a learning environment based on Knowledge Graphs".

- Nicholas Halliwell on "Evaluating and improving explanation quality of graph neural network link prediction on knowledge graph" [91]

- Ahmed El Amine Djebri on "Uncertainty Management for Linked Data Reliability on the Semantic Web" [90].

5.3   Awards

Best Highlight Paper at IC-PFIA 2022 for Lucie Cadorel [50].

6   New software and platforms

6.1   New software

6.1.1   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.

URL: http://ns.inria.fr/acta/

Contact: Serena Villata

6.1.2 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 -> 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 -> 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.

URL: http://covid19.i3s.unice.fr:8080/arviz/

Publication: hal-03292140

Contact: Marco Alba Winckler

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

6.1.3 CORESE

Name: COnceptual REsource Search Engine

Keywords: Semantic Web, Search Engine, RDF, SPARQL

Functional Description: Corese is a Semantic Web Factory, it implements W3C RDF, RDFS, OWL RL, SHACL, SPARQL 1 .1 Query and Update as well as RDF Inference Rules.

Furthermore, Corese query language integrates original features such as approximate search and extended Property Path. It provides STTL: SPARQL Template Transformation Language for RDF graphs. It also provides LDScript: a Script Language for Linked Data. Corese provides distributed federated query processing.

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

Contact: Olivier Corby

Participants: Erwan Demairy, Fabien Gandon, Fuqi Song, Olivier Corby, Olivier Savoie, Virginie Bottollier

Partners: I3S, Mnemotix

6.1.4   Corese Server

Name: Corese Server

Keywords: Semantic Web, RDF, SPARQL, OWL, SHACL

Scientific Description: This library provides a Web server to interact with Corese via HTTP requests. In includes a SPARQL endpoint and the STTL display engine to generate portals from linked data (RDF).

Functional Description: This library provides a Web server to interact with Corese via HTTP requests. In includes a SPARQL endpoint and the STTL display engine to generate portals from linked data (RDF).

Contact: Olivier Corby

Participants: Alban Gaignard, Fuqi Song, Olivier Corby

Partner: I3S

6.1.5   CREEP semantic technology

Keywords: Natural language processing, Machine learning, Artificial intelligence

Scientific Description: The software provides a modular architecture specifically tailored at the classification of cyberbullying and offensive content on social media platforms. The system can use a variety of features (ngrams, different word embeddings, etc) and all the netwok parameters (number of hidden layers, dropout, etc) can be altered by using a configuration file.

Functional Description: The software uses machine learning techniques to classify cyberbullying instances in social media interactions.

Release Contributions: Attention mechanism, Hyperparameters for emoji in config file, Predictions output, Streamlined labeling of arbitrary files

Publications: hal-01906096v1, hal-01920266v1

Contact: Michele Corazza

Participants: Michele Corazza, Elena Cabrio, Serena Villata

6.1.6   DBpedia

Name: DBpedia

Keywords: RDF, 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

6.1.7 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

6.1.8 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.

URL: https://github.com/Wimmics/dekalog

Contact: Pierre Maillot

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

6.1.9 ISSA-pipeline

Name: Processing pipeline of the ISSA project

Keywords: Indexation, Semantic Web, Natural language processing, Knowledge graph, Open Access, Open data, LOD - Linked open data

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

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

Contact: Franck Michel

Partners: CIRAD, IMT Mines Alès

6.1.10 KartoGraphI

Keywords: Semantic Web, LOD - Linked open data

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

6.1.11   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 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 presented by Governatori et al. 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., Licensius5 or Creative Commons Choose service6). 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

6.1.12   SPARQL Micro-services

Name: SPARQL micro-services

Keywords: Web API, SPARQL, Microservices, LOD - Linked open data, Data integration

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.

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

Publications: hal-02060966, hal-01722792, hal-01947589, hal-02168164

Author: Franck Michel

Contact: Franck Michel

6.1.13 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

6.1.14 MGExplorer

Name: Multivariate Graph Explorer

Keyword: Information visualization

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 loosing information on previous analyses.

Functional Description: MGExplorer is an information visualization tool suite that integrates many information visualization techniques aimed at supporting the exploration of multivariate graphs. MGExplorer allows users to choose and combine the information visualization techniques creating a graph that describes the exploratory path of dataset. It is an application based on the D3.JS library, which is executable in a web browser. The use of MGExplorer requires a customization to connect the dashboard to a SPARQL endpoint. MGExplorer has been customized to facilitate the search of scientific articles related to covid.

Release Contributions: Visualization of data extracted from linked data datasets.

URL: http://covid19.i3s.unice.fr:8080/

Publications: hal-03292172, hal-03404572, hal-03404580

Contact: Marco Alba Winckler

Participants: Aline Menin, Marco Alba Winckler, Olivier Corby

Partner: Universidade Federal do Rio Grande do Sul

6.1.15   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

Author: Franck Michel

Contact: Franck Michel

6.1.16   Muvin

Name: Multimodal Visualization of Networks

Keywords: Data visualization, Music, LOD - Linked open data

Functional Description: Muvin supports the exploration of a two-layer network describing the collaborations between artists and the discography of an artist, defined by the albums and songs released by the artist over time. It implements an incremental approach, allowing the user to dynamically import data from a SPARQL endpoint to the exploration flow. Furthermore, this approach seeks to improve user perception by associating audio to the visualization, in a way that the users can listen to the songs visually represented in their screen.

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

Contact: Aline Menin

6.1.17   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

Keyword: Tube guitar amplifier simulator for web browser

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.

News of the Year: Best paper at WebAudio Conference 2020.

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

Partner: Amp Track Ltd, Finland

7   New results

7.1   User Modeling and Designing Interaction

7.1.1   Incremental visual exploration of linked data

> **Participants:**   Marco Winckler, Aline Menin, Olivier Corby, Catherine Faron, Alain Giboin.

Information visualization techniques are useful to discover patterns and causal relationships within LOD datasets. However, since the discovery process is often exploratory (i.e., users have no predefined goal and do not expect a particular outcome), when users find something interesting, they should be able to (i) retrace their exploratory path to explain how results have been found, and (ii) branch out the exploratory path to compare data observed in different views or found in different datasets. Furthermore, as most of LOD datasets are very specialized, users often need to explore multiple datasets to obtain the knowledge required to support decision-making processes. Thus, the design of visualization tools is confronted with two main challenges: the visualization system should provide multiple views to enable the exploration of different or complementary perspectives to the data; and the system should support the combination of diverse data sources during the exploration process. To our knowledge, the existing tools before our work, are limited to visualizing a single dataset at a time and, often, use static and preprocessed data. Thus, we proposed the concept of follow-up queries to allow users to create queries on demand during the exploratory process while connecting multiple LOD datasets with chained views. Our approach relies on a exploration process supported by the use of predefined SPARQL queries that the user can select on-the-fly to retrieve data from different SPARQL endpoints. It enables users to enrich the ongoing analysis by bringing external and complementary data to the exploration process, while also supporting the visual analysis and comparison of different subsets of data (from the same or different SPARQL endpoints) and, thus, the incremental exploration of the LOD cloud. The resulting publication [36] presents a generic visualization approach to assist the analysis of multiple LOD datasets based on the concepts of chained views and follow-up queries. We demonstrate the feasibility of our approach via four use case scenarios and a formative evaluation where we explore scholarly data described by RDF graphs publicly available through SPARQL endpoints. These scenarios demonstrate how the tool supports (i) composing, running, and visualizing the results of a query; (ii) subsetting the data and exploring it via different visualization techniques; (iii) instantiating a follow-up query to retrieve external data; and (iv) querying a different database and compare datasets. The usability and usefulness of the proposed approach is confirmed by results obtained with a series of semi-structured interviews. The results are encouraging while showing the relevance of the approach to explore big linked data. This work resulted in a visualization tool, called LDViz, publicly accessible at dataviz.i3s.unice.fr/ldviz. The source code is also open and published as 10.5281/zenodo.6511782. A large study concerning the scalability of the tools has demonstrated over 420 public end points [89].

7.1.2 Interaction with extended reality

Participants: Aline Menin, Clément Quere, Florent Robert, Hui-Yin Wu, Marco Winckler.

Virtual reality (VR) and Augmented Reality (AR) offer extraordinary opportunities in user behavior research to study and observe how people interact in immersive 3D environments and in situations of the so-called extended reality. A major challenge of designing these 3D experiences and user tasks, however, lies in bridging the inter-relational gaps of perception between the designer, the user, and the 3D scene. In the context of the PhD thesis of Florent Robert [78] we have started a series of studies aiming to understand how the design of user of 3D scenes affect the user perception and how such perception affect decision-making processes. Our ultimate goals is try to understand how the many components of the user interaction (including user attention) might affect the embodided experience virtual immersive environments [62][63]. Fort that, we have proposed a tool suppport [41], called 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. This framework allows to describe 3D scenes that embed objects with semantic content and tasks to be performed in the environment. The embodied user experience is captured by a set of physiological sensors and the results (in terms of user attention and emotions inferred from physiological data) can be traced to objects and tasks modeled in the GUST-3D framework. We want to extend the study along the PhD thesis of Clement Quere, started in September 2022, for use of VR/AR technologies to visualize large volumes of data and to observe the improvement of users' abilities to understand the relationships between data (patterns, trends, correlations, etc.) and to make reasoned decisions. On the one hand, this thesis hypothesizes that the use of hand and wrist movements activates the proprioceptive sense of the users to enhance the learning and memorization processes, so it aims to identify interaction techniques in 3D space that are natural for the user. On the other hand, we also explore the hypothesis that diverse expertise favors the resolution of analytical problems (complex, uncertain, and potentially ill-defined) and we propose the definition of a method based on geocollaboration to explore the data. The expected results are: (1) the understanding of the uses and interactions for spatio-temporal data mining in an immersive environment, (2) a generic, flexible and extensible Framework capable of displaying several visualization techniques and (3) fine analysis of spatio-temporal data in the application domains of two case studies: academic mobility and urban mobility.

7.1.3 Use of annotations for interactive support of decision-making processes

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

The amount and complexity of digital data has increased exponentially during the last couple of decades. These data contain valuable information to support decision-making processes in several application domains. In our research we consider data in their many possible incarnations, including digital data used to create interactive systems (such as database models, prototypes, models describing the systems architecture, etc.) and digital documents (such as Web pages, endpoints, etc.). To start, we investigate the fundamental role played by annotations to support decision-making processes. This initiative is grounded on the fact that many relevant information (such as rational about the design, decisions made, recommendations, etc.) are not explicitly captured and represented in the models/artefacts that are commonly used to describe interactive systems. We have identified a set of problems related to the way annotations are created and used along the development process of interactive systems. For that, we proposed a model-based approach that was conceived to handle annotations in a systematic way along the development process [40]. As part of the solution, we proposed an annotation model built upon the W3C's Web Annotation Data Model. Some preliminary results suggest that our approach could be generalized to annotate the exploratory processes using open source datasets. In this respect, we identified three potential uses of annotations: (i) documenting findings (including errors in the dataset),

(ii) supporting collaborative reasoning among teammates, and (iii) analysing provenance during the exploratory process. Therefore, we investigate the use of annotations during the visual exploration of datasets assisted by chained visualization techniques. The preliminary results have been published in an international conference [79] and we are working on the generalization of the approach and the corresponding tools.

7.1.4 Incremental and multimodal visualization of music artists' discographies and collaborations

Participants: Aline Menin, Michel Buffa, Maroua Tikat, Benjamin Molinet, Guillaume Pellerin, Laurent Pottier, Franck Michel, Marco Winckler.

In this work, we support musicologists on the analysis of music data. Nowadays, there are various and multidimensional music datasets available on the Web describing musical content over aspects such as lyrics, chords, sounds, metadata, etc. In this context, visualization techniques are suitable tools to facilitate the access to the data, while being capable of highlighting relationships between structural elements of music. Particularly, we focus on assisting the exploration and analysis of artists' discographies and collaborations. Artist collaborations in music tend to result in successful songs, which analysis can help to understand the impact of these collaborative projects on artists' careers and their music style. Further to studying aspects of cultural context (e.g., release dates, collaborations, locations), it is essential for musicologists to analyze the timbre and annotations of the audio signal, as well as careful listening to the songs, when studying the acoustic characteristics of songs. Thus, we also explore the potential of the audio dimension to further support musicologists on their analysis, as well as to improve user perception through an auditory data exploration approach. In the resulting publication [72], we present our approach through a case of study of the visualization of data from the WASABI RDF knowledge graph, which gathers metadata describing over two million commercial songs (200K albums and 77K artists – mainly from pop/rock culture). It includes metadata about songs, albums, and artists (e.g., artists, discography, producers, dates, etc.) retrieved from multiple data sources on the Web. The WASABI dataset contains data on commercial music recordings from 1922 to 2022, but it does not provide explicit information about collaborations between artists and how they intersect over time. The proposed visualization approach supports the exploration of a social network of artists and bands derived from their recordings and mapped to the dimensions of *association* (type of contribution) and *time* (progression of the artist's career). In particular, our contributions are a Web-based interactive tool to visualize the discography of artists / groups and collaborations between musical artists and groups across time; an auditory exploration approach based on audio thumbnailing to support user perception throughout the visual exploration process; a direct link with external services that support further exploration of songs through MIR analysis. This work resulted in a visualization tool, called Muvin 6.1.16, publicly accessible at dataviz.i3s.unice.fr/muvin/.

7.1.5 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 [31] of the audio files provided by Deezer. This dataset is still exploited by actual projects inside the team such as the previous presented work on "Incremental and multimodal visualization of music artists, discographies and collaborations". Other initiatives closely related to the WASABI datasets include several Web Audio interactive applications and frameworks. Web Audio Modules 2 is a WebAudio plugin standard for developing high performance plugins in the browser [46, 44]. We also developed new methods for real-time tube guitar amplifier simulations that run in the browser [47, 67]. Some of these results are unique in the world as in 2022, 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 [37, 77, 48].

7.1.6 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. [95]

7.2 Communities and Social Interactions Analysis

7.2.1 Autonomous agents in a social and ubiquitous Web

Participants: Andrei Ciortea, Olivier Corby, Fabien Gandon, Franck Michel.

Recent W3C recommendations for the Web of Things (WoT) and the Social Web are turning hypermedia into a homogeneous information fabric that interconnects heterogeneous resources: devices, people, information resources, abstract concepts, etc. The integration of multi-agent systems with such hypermedia environments now provides a means to distribute autonomous behavior in worldwide pervasive systems. A central problem then is to enable autonomous agents to discover heterogeneous resources in world wide and dynamic hypermedia environments. This is a problem in particular in WoT environments that rely on open standards and evolve rapidly—thus requiring agents to adapt their behavior at runtime in pursuit of their design objectives. To this end, we developed a hypermedia search engine for the WoT that allows autonomous agents to perform approximate search queries in order to retrieve relevant resources in their environment in (weak) real time. The search engine crawls dynamic WoT environments to discover and index device metadata described with the W3C WoT Thing Description, and exposes a SPARQL endpoint that agents can use for approximate search. To demonstrate the feasibility of our approach, we implemented a prototype application for the maintenance of industrial robots in worldwide manufacturing systems. The prototype demonstrates that our semantic hypermedia search engine enhances the flexibility and agility of autonomous agents in a social and ubiquitous Web [100].

7.2.2 Abusive language detection

Participants: Elena Cabrio, Serena Villata, Anais Ollagnier.

Recent studies have highlighted the importance to reach a fine-grained online hate speech characterisation to provide appropriate solutions to curb online abusive behaviours. In this direction, we proposed a full pipeline that enables to capture targeting characteristics in hatred contents (i.e., types of hate, such as race and religion) aiming at improving the understanding on how hate is conveyed on Twitter. Our contribution is threefold: (1) we leverage multiple data views of a different nature to contrast different kinds of abusive behaviours expressed towards targets; (2) we develop a full pipeline relying on a multi-view clustering technique to address the task of hate speech target characterisation; and (3) we propose a methodology to assess the quality of generated hate speech target communities. Relying on multiple data views built from multilingual pre-trained language models (i.e., multilingual BERT and

multilingual Universal Sentence Encoder) and the Multi-view Spectral Clustering (MvSC) algorithm, the experiments conducted on a freely available multilingual dataset of tweets (i.e., the MLMA hate speech dataset) show that most of the configurations of the proposed pipeline significantly outperforms state-of-the-art clustering algorithms on all the tested clustering quality metrics on both French and English. A journal paper is undersubmission on the results we obtained on this topic.

In addition, we carried out a data collection in the context of the UCA IDEX OTESIA project "Artificial Intelligence to prevent cyberviolence, cyberbullying and hate speech online." [1] to creat a dataset of aggressive chats in French collected through a role-playing game in high-schools. The collected conversations have been annotated with the participant roles (victim, bully, and bystanders), the presence of hate speech (content that mocks, insults, or discriminates against a person or group based on specific characteristics such as colour, ethnicity, gender, sexual orientation, nationality, religion, or other characteristics) and whether utterances use different humour figurative devices such as sarcasm or irony. Moreover, we have also introduced a new annotation layer referring to the different types of verbal abuse present in the message, defined as a type of psychological/mental abuse that involves the use of written language relying on derogatory terms, the delivery of statements intended to demean, humiliate, blame or threaten the victim with the aim of decreasing its self-confidence and making it feel powerless. The identification of the different types of verbal abuse will allow to investigate (and learn, from a computational perspective) strategies used by cyberhate perpetrators to cause emotional harm, and to reach insights about how victims respond to bullying/victimisation. In particular, this activity resulted in a new dataset called CyberAgressionAdo-V1 [76], containing aggressive multiparty chats in French collected through a role-playing game in high-schools, and annotated at these different layers. In this dataset, we analysed the different types of aggression and verbal abuse depending on the targeted victims (individuals or communities) emerging from the collected data.

7.2.3   Fallacious Argument Classification in Political Debates

Participants: Pierpaolo Goffredo, Shohreh Haddadan, Vorakit Vorakitphan, Elena Cabrio, Serena Villata.

Fallacies play a prominent role in argumentation since antiquity due to their contribution to argumentation in critical thinking education. Their role is even more crucial nowadays as contemporary argumentation technologies face challenging tasks as misleading and manipulative information detection in news articles and political discourse, and counter-narrative generation. Despite some work in this direction, the issue of classifying arguments as being fallacious largely remains a challenging and an unsolved task. Our contribution is twofold: first, we present a novel annotated resource of 31 political debates from the U.S. Presidential Campaigns, where we annotated six main categories of fallacious arguments (i.e., ad hominem, appeal to authority, appeal to emotion, false cause, slogan, slippery slope) leading to 1628 annotated fallacious arguments; second, we tackle this novel task of fallacious argument classification and we define a neural architecture based on transformers outperforming state-of-the-art results and standard baselines. Our results show the important role played by argument components and relations in this task [60, 81].

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

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

7.3.1   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. The collaboration with the "Muséum National d'Histoire Naturelle" of Paris (MNHN) goes on along several axes.

Since 2019 the MNHN has been using our SPARQL Micro-Services architecture and framework to help biologists in editing taxonomic information by confronting multiple, heterogeneous data sources. This collaboration is going on and the MNHN now heavily relies on those services for daily activities. Furthermore, at the occasion of multiple meetings thoughout the year, we have been accompanying the MNHN in the modelling a a biodiversity thesaurus, and the addition of life traits to this thesaurus, such that this will be compatible with and usable by multiple groups within the museum.

7.3.2   The WASABI Song Corpus with Lyrics Annotations.

Participants: Elena Cabrio, Michael Fell, Michel Buffa.

The WASABI Song Corpus is a large corpus of songs enriched with metadata extracted from music databases on the Web, and resulting from the processing of song lyrics and from audio analysis. Given that lyrics encode an important part of the semantics of a song, we have focused on the design and application of methods to extract relevant information from the lyrics, such as their structure segmentation, their topics, the explicitness of the lyrics content, the salient passages of a song and the emotions conveyed. So far, the corpus contains 1.73M songs with lyrics (1.41M unique lyrics) annotated at different levels with the output of the above mentioned methods. Such corpus labels and the provided methods can be exploited by music search engines and music professionals (e.g. journalists, radio presenters) to better handle large collections of lyrics, allowing an intelligent browsing, categorization and recommendation of songs [31].

7.3.3   Evolutionary agent-based evaluation of the sustainability of different knowledge sharing strategies in open multi-agent systems

Participants: Stefan Sarkadi, Fabien Gandon.

The advancement of agent technologies and their deployment in various fields of application has brought numerous benefits w.r.t. knowledge or data gathering and processing. However, one of the key challenges in deploying artificial intelligent agents in an open environment like the Web is their interoperability. Even tough research and development of agent technologies on the Semantic Web has advanced significantly, artificial agents live on the Web in silos, that is in very limited domains, isolated from other systems and agents that live on the Web. In this work we setup a simulation framework and evaluation based on evolutionary agent-based modeling to empirically test how sustainable different strategies are for knowledge sharing in open multi-agent systems and to see which of these strategies could actually enable global interoperability between Web agents. The first results are showing the interest of translation-based approaches and the need for further incentives to support these [38].

7.3.4   Agent-Based Modeling of Economical Systems

Participants: Andrea Tettamanzi.

A central question in economics is how a society accepts money, defined as a commodity used as a medium of exchange, as an unplanned outcome of the individual interactions. Together with economists of the SKEMA Business School, we investigated the reasons for the failure of previous work to have boundedly rational agents learn speculative strategies that were theoretically predicted. We started with an agent-based model proposed in the literature, where the intelligence of the agents is guided by a

learning classifier system that is shown to be capable of learning trade strategies (core strategies) that involve short sequences of trades. We tested several modifications of the original model and we came up with a set of assumptions that enable the spontaneous emergence of speculative strategies, which explain the emergence of money even when the agents have bounded rationality [55].

7.3.5 LOV-ES: Guiding the Ontology Selection to Structure Textual Data using Topic Modeling

Participants: Damien Graux, Anaïs Ollagnier.

On-line availability of text corpora nowadays allow data practitioners to build complex knowledge combining various sources. One common shared challenge relies in the modelisation of intermediate knowledge structures able to gather at once the various topics present in the texts. Practically, practitioners often go through the creation of vocabularies. In order to help these domain experts, we design LOVES: a solution able to help them in this creative process, guiding them in the selection and the combination of already existing vocabularies available online. Technically, our solution relies on LDA to detect topics and on the LOV to then propose candidate vocabularies. [61].

7.3.6 W3C Data activity

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. Two new groups were created in 2022, where Wimmics members are largely involved. RDF Canonicalization and Hash aims at providing a way to hash RDF data independently of the way they are represented. RDF-star is chartered to publish the new version of RDF and SPARQL, extending them with the ability to make statements about statements.

Finally, work has started towards the creation of a working group 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.

7.4 Analyzing and Reasoning on Heterogeneous Semantic Graphs

7.4.1 Uncertainty Evaluation for Linked Data

Participants: Ahmed Elamine Djebri, Fabien Gandon, Andrea Tettamanzi.

For data sources to ensure providing reliable linked data, they need to indicate information about the (un)certainty of their data based on the views of their consumers. In addition, uncertainty information in terms of Semantic Web has also to be encoded into a readable, publishable, and exchangeable format to increase the interoperability of systems. We introduced a novel approach to evaluate the uncertainty of data in an RDF dataset based on its links with other datasets. We proposed to evaluate uncertainty for sets of statements related to user-selected resources by exploiting their similarity interlinks with external resources. Our data-driven approach translates each interlink into a set of links referring to the position of a target dataset from a reference dataset, based on both object and predicate similarities. We showed

how our approach can be implemented and present an evaluation with real-world datasets. Finally, we discussed updating the publishable uncertainty values. Details are available online. This work was defended in the PhD thesis of Ahmed Elamine Djebri [90].

In collaboration with colleagues of IRIT, we investigated how metadata about the uncertainty of knowledge contained in a knowledge base can be expressed parsimoniously and used for reasoning. We proposed an approach based on possibility theory, whereby a classical knowledge base plus metadata about the degree of validity and completeness of some of its portions are used to represent a possibilistic belief base. We show how reasoning on such belief base can be carried out by using a classical reasoner [53].

7.4.2 A Semantic Model for Meteorological Knowledge Graphs

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

The great interest advocated by the agronomy and biodiversity communities in the development of crop models coupled to weather and climate models has led to the need for datasets of meteorological observations in which data are semantically described and integrated. For this purpose, in the context of the D2KAB ANR project, we proposed a semantic model to represent and publish meteorological observational data as Linked Data. Our model reuses a network of existing ontologies to capture the semantics of data, it covers multiple dimensions of meteorological data including geospatial, temporal, observational, and provenance characteristics. Our proposition also provides a SKOS vocabulary of terms to describe domain-specific observable properties and features. We paid specific attention to propose a model that adheres to LD best practices and standards, thereby allowing for its extension and re-use by several meteorological data producers, and making it capable of accommodating multiple application domains. This work has been published and presented at ICWE 2022 [83] and ESWC 2022 [96].

Based on this semantic model, we built the WeKG-MF knowledge graph considering the open weather observations published by Météo-France. The SPARQL WeKG-MF[2] endpoint allows users to retrieve weather observations recorded every 3 hours by different sensors hosted by weather stations and related to different parameters (air temperature, humidity, wind speed, precipitation, atmospheric pressure, etc.). In order to enable an interactive exploration of the WeKG-MF graph, we released a Web application [3] that enables lay users to visualize weather observational data at different levels of spatio-temporal granularity, and hence, it offers multi-level 'tours' based on high-level aggregated views together with on-demand fine-grained data, and this through a unified multi-visualisations interface. This work has been published and presented in [82].

In addition we worked on the alignment of two complementary knowledge graphs useful in the agricultural domain: the crop usage thesaurus (CUT) and the French national taxonomic register TAXREF for fauna, flora and fungi. Several alignment methods specific to this use case were implemented. The results show that in this domain it will be necessary to clean up the automatically generated alignments [73].

7.4.3 Wheat-KG: a Knowledge graph for wheat genomics studies

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

One of the main challenges in wheat genomics is the large size and complexity of the wheat genome. Research in wheat genomics has already led to several important advances, such as the development of wheat varieties with enhanced disease resistance and improved nutritional content. Indeed, experts in wheat genomics are always interested in deeper understanding of the genotype-phenotype relationships. Harvesting scientific literature may support them to understand hidden interactions between genomic

---

[2]The SPARQL endpoint is accessible online
[3]A demo of the first release of the Web application is accessible online

entities based on their co-occurrence in scientific publications. In the context of the D2KAB project, we propose to structure and integrate semantic annotations extracted automatically using NLP tools into a knowledge graph. The main purpose is to bridge the gap between scientific results presented in publications and experts needs to explore literature and find answers to complex questions. This research work has been presented in a national conference [87].

7.4.4 Corese Semantic Web Factory

Participants: Rémi Ceres, Olivier Corby.

Corese is an open source Semantic Web platform that implements W3C languages such as RDF, RDFS, OWL RL, SHACL, SPARQL and extensions such as SPARQL Function, SPARQL Transformation and SPARQL Rule.

In the context of the National research program in artificial intelligence (PNRIA) [4] and in collaboration with the Mnemotix cooperative [5], we continued to work on the industrialization of Corese.

To improve the distribution of Corese, we pushed the project to Maven Central [6] and created a Docker image for Corese-Server [7]. We also completed the documentation of Corese [8]. We updated the Jetty version in Corese-Server.

In addition to the Corese library, Corese-GUI, and Corese-Server, we created two new interfaces: Corese-Python and Corese-CLI. Corese-CLI allows users to use Corese through the command line, including converting an RDF file between different serialization formats, running a SPARQL query, checking OWL profiles, and running an LDSCRIPT file. Corese-Python allows users to use the Corese library through Python code.

We implemented persistence in Corese by improving the DataManager API, which allows the Corese SPARQL engine to connect with external storage systems. We implemented new data managers for the RDF4J model, Jena TDB1, InteGrall Storage system, Corese Graph, JSON, and XML. This allows users to connect the SPARQL engine of Corese with external storage systems and store and retrieve data in various formats.

Corese was extended with RDF-star and SPARQL-star [9] that enable users to annotate RDF triples with RDF triples.

Corese federated query engine has been extended to handle properly join in basic graph patterns. Heuristics have been designed in order to group triples in appropriate connected graph patterns with respect to remote endpoints. The federated query engine is also able to process source indentification in a Semantic Web graph index such as the one designed by the team in the Dekalog ANR project.

Corese core functionnalities such as SPARQL, SPARQL Function (LDScript), SPARQL Transformation (STTL), SPARQL Rule, approximate search, etc. have been ported on the data manager generic broker to external storage. Hence, Corese functionnalities are available with external storage such as Jena TDB1 and RDF4J for which data managers have been implemented.

Web sites: Corese Web site, Corese github URL.

7.4.5 SHACL Extension

Participants: Olivier Corby, Iliana Petrova, Fabien Gandon, Catherine Faron.

In the context of a collaboration with Stanford University, we worked on extensions of W3C SHACL Shape Constraint Language [10].

---

[4] Web Site PNRIA
[5] Web Site Mnemotix
[6] Corese Web page on Maven Central
[7] Docker-hub Web page of Corese-Server
[8] Documentation of Corese
[9] RDF-star
[10] see SHACL Recommendation

We conducted a study on large, active, and recognized ontology projects (ex. Gene Ontology, Human Phenotype Ontology, Mondo Disease Ontology, Ontology for Biomedical Investigations, OBO Foundry, etc.) as well as an analysis of several existing tools, methodologies and guidelines for ontological engineering.

As a result we identified several sets of ontology validation constraints that fall into six big clusters: **i)** formalization/modeling checks; **ii)** terminological/writing checks; **iii)** documentation/ editorial practices, terminology-level checks; **iv)** coherence between terminology and formalization; **v)** metamodel-based checks; **vi)** integration/interoperability/data checking. These can be further refined depending on whether they are specific to RDFS/OWL meta-model, domain/ontology specific, or Linked Data specific. This precise categorization of the ontology validation constraints allowed us to analyse the needs and impact of the extension we are targeting in terms of semantic expressiveness, computational complexity of the validation, and current syntax of the SHACL language.

We then concentrated on the formalization of the semantic extensions and their validation methods and came up with a proposal of a corresponding syntactic extensions of SHACL.

The formal specification of the identified extensions enabled us to proceed with the implementation of a prototype plugin for Protégé (Stanford's widely used ontology editor) based on the Corese engine and which extends the SHACL standard with these newly proposed capabilities.

7.4.6 Learning object recommendation based on learning objectives

Participants: Molka Dhouib, Catherine Faron, Oscar Rodríguez Rocha.

With the digital transformation, adaptation and development of skills have become major factors in improving employee and business performance. Understanding the needs of employees and helping them achieve their career development goals is a real challenge today. We implement an automatic recommendation system that allows learners to find relevant learning objects based on their learning goals. This matching task is mainly based on determining the semantic similarity between the goals and the textual content of the learning objects. We comparatively evaluated three state-of-the-art pre-trained sentence embedding models for the learning object recommendation task, and we studied the impact of adapting and using an existing repository versus building an internal repository to define learning objectives. Experimental results show that the use of these sentence embedding models in the recommendation process outperforms the Elasticsearch BM25 model classically used in the industry, and the use of an internal repository improves the recommendation result compared to the adaptation of an existing standard repository. The result of this work is presented in [85] and [86].

7.4.7 Identifying argumentative structures in clinical trials

Participants: Elena Cabrio, Serena Villata, Santiago Marro, Benjamin Molinet.

In the latest years, the healthcare domain has seen an increasing interest in the definition of *intelligent* systems to support clinicians in their everyday tasks and activities. Among others, the field of Evidence-Based Medicine is impacted by this twist, with the aim to combine the reasoning frameworks proposed thus far in the field with mining algorithms to extract structured information from clinical trials, clinical guidelines, and Electronic Health Records. In this work, we go beyond the state of the art by proposing a new end-to-end pipeline to address argumentative outcome analysis on clinical trials. More precisely, our pipeline is composed of *(i)* an Argument Mining module to extract and classify argumentative components (i.e., evidence and claims of the trial) and their relations (i.e., support, attack), and *(ii)* an outcome analysis module to identify and classify the effects (i.e., improved, increased, decreased, no difference, no occurrence) of an intervention on the outcome of the trial, based on PICO elements. We annotated a dataset composed of more than 500 abstracts of Randomized Controlled Trials (RCT) from the MEDLINE database, leading to a labeled dataset with 4198 argument components, 2601 argument

relations, and 3351 outcomes on five different diseases (i.e., *neoplasm, glaucoma, hepatitis, diabetes, hypertension*). We experiment with deep bidirectional transformers in combination with different neural architectures (i.e., LSTM, GRU and CRF) and obtain a macro F1-score of .87 for component detection and .68 for relation prediction, outperforming current state-of-the-art end-to-end Argument Mining systems, and a macro F1-score of .80 for outcome classification.

In this context, we also released a new version of the ACTA demo. ACTA 2.0[11] is an automated tool which relies on Argument Mining methods to analyse the abstracts of clinical trials to extract argument components and relations to support evidence-based clinical decision making. ACTA 2.0 allows also for the identification of PICO (Patient, Intervention, Comparison, Outcome) elements, and the analysis of the effects of an intervention on the outcomes of the study. A REST API is also provided to exploit the tool's functionalities [75].

7.4.8 Qualitative evaluation of arguments in persuasive essais

Participants: Elena Cabrio, Serena Villata, Santiago Marro.

Argumentation is used by people both internally, by evaluating arguments and counterarguments to make sense of a situation and take a decision, and externally, e.g., in a debate, by exchanging arguments to reach an agreement or to promote an individual position. In this context, the assessment of the quality of the arguments is of extreme importance, as it strongly influences the evaluation of the overall argumentation, impacting on the decision making process. The automatic assessment of the quality of natural language arguments is recently attracting interest in the Argument Mining field. However, the issue of automatically assessing the quality of an argumentation largely remains a challenging unsolved task.

Our contribution is twofold: first, we present a novel resource of 402 student persuasive essays, where three main quality dimensions (i.e., *cogency, rhetoric*, and *reasonableness*) have been annotated, leading to 1908 arguments tagged with quality facets [69]; second, we address this novel task of argumentation quality assessment proposing a novel neural architecture based on graph embeddings, that combines both the textual features of the natural language arguments and the overall argument graph, i.e., considering also the support and attack relations holding among the arguments. Results on the persuasive essays dataset outperform state-of-the-art and standard baselines' performance [70].

7.4.9 Identification of the information captured by Knowledge Graph Embeddings

Participants: Antonia Ettorre, Anna Bobasheva, Catherine Faron, Franck Michel.

The recent growth in the utilization of Knowledge Graphs has been powered by the expanding landscape of Graph Embedding techniques, which facilitates the manipulation of the vast and sparse information described by such Knowledge Graphs. Although the effectiveness of Knowledge Graph Embeddings has been proved on many occasions and for many contexts, the interpretability of such vector representations remains an open issue. To tackle it, we provided a systematic approach to decode and make sense of the knowledge captured by Graph Embeddings. We proposed a tool called Stunning Doodle to visualize jointly a graph and the embeddings of the nodes thereof [57, 56], and verify whether Graph Embeddings are able to encode certain properties of the graph elements they represent. We also showed that using purely graph-based techniques such as the link prediction is able to obtain performances in terms of pedagogical resources recommendation, that are as effective as common knowledge tracing approaches [93].

7.4.10 RDF Mining

---

[11]ACTA 2.0 online demo.

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

An optimization on assessment of OWL SubClassOf axioms against an RDF knowledge graph [58] has been proposed to respond to a high computational cost, especially in terms of computation time (CPU).

On the other hand, our evolutionary approach critically relies on (candidate) axiom scoring. In practice, testing an axiom boils down to computing an acceptability score, measuring the extent to which the axiom is compatible with the recorded facts. Methods to approximate the semantics of given types of axioms have been thoroughly investigated in the last decade, but a promising alternative to their direct computation is to train a surrogate model on a sample of candidate axioms for which the score is already available, to learn to predict the score of a novel, unseen candidate axiom. This is the main objective Ali Ballout's thesis, whose first results were a method to predict the score of atomic OWL axioms based on axiom similarity [43]. Insights on why axiom similarity works well for this type of task has been offered by an investigation on a sort of toy problem related to axiom scoring, namely using machine learning methods to guess the truth value of propositional formulas based on a measure of semantic similarity [42].

7.4.11   Capturing Geospatial Knowledge from Real-Estate Advertisements

Participants: Lucie Cadorel, Andrea Tettamanzi.

In the framework of a CIFRE thesis with Kinaxia, we have proposed a workflow to extract geographic and spatial entities based on a BiLSTM-CRF architecture with a concatenation of several text representations and to extract spatial relations, to build a structured Geospatial knowledge base. This pipeline has been applied it to the case of French housing advertisements, which generally provide information about a property's location and neighbourhood. 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 [50, 51, 52].

7.4.12   ISSA: semantic indexing of scientific articles and advanced services

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

Faced with the ever-increasing number of scientific publications, in October 2020 we started ISSA project with the following goals: (1) provide a generic, reusable and extensible pipeline for the analysis and processing of articles of an open scientific archive, (2) translate the result into a semantic index stored and represented as an RDF knowledge graph; (3) develop innovative search and visualization services that leverage this index to allow researchers, decision makers or scientific information professionals to explore thematic association rules, networks of co-publications, articles with co-occurring topics, etc.

The project ended in December 2022, and we could deliver a pipeline for the semantic indexing of the scientific publications from Agritrop, Cirad's scientific archive of 110,000+ resources among which 12,000 open access articles. Indexing is performed on two levels: thematic and geographic descriptors characterizing the article as a whole, and named entities extracted from the articles' text. Descriptors and named entities are linked with reference knowlegde bases such as Wikidata, DBpedia, Geonames, and Agrovoc. We have also prototyped visualization services that render the entities and descriptors extracted during the indexing process and allow exploring the graph of authors, institutions, research topics, etc.

The pipeline as well as visuzalization tools are DOI-identified and available under an open license on public repositories. The outcome of this pipeline is a knowledge graph that we also made public as a downloadable dump (DOI: 10.5281/zenodo.6505847) and through a public SPARQL endpoint. We published the results of the project in two international conferences: [80] and [74].

7.4.13 IndeGx: A Model and a Framework for Indexing Linked Datasets and their Knowledge Graphs with SPARQL-based Test Suits

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

The joint exploitation of RDF datasets relies on the knowledge of their content, of their endpoints, and of what they have in common. Yet, not every dataset contains a self-description, and not every endpoint can handle the complex queries used to generate such a description.

As part of the ANR DeKaloG, we proposed a standard-based approach to generate the description of a dataset. The description generated as well as the process of its computation are expressed using standard vocabularies and languages. We have implemented our approach into a framework, called IndeGx, to automatically generate the description of datasets and endpoints and collect them in an index. We have experimented IndeGx on a set of 339 active knowledge bases.

Several visualisations were also generated from IndeGx and are available online: IndeGx Web Site.

7.4.14 Metadatamatic: A tool for the description of RDF knowledge bases

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

During the experimentations done as part of the development of IndeGx, we observed that less than 10% of accessible RDF datasets offer a description of their provenance. We theorize that this is partly due to the difficulty for data providers to learn how to create such descriptions. Some initiatives such as the COST KG working group, or a European Commission action have proposed guides consisting of lists of mandatory and recommended classes and properties to use.

We propose Metadatamatic, an online tool based on the before-mentioned initiatives to generate a KB description. Metadatamatic aims to bypass this learning problem and help further the description of a KB. Metadatamatic generates a description using well-established vocabulary from a simple Web form. It also offers to extract some parts of the description, such as the list of vocabularies used in the data, automatically from the content of the KB. We hope that this tool will lead to an improvment of the usability of Linked Data in general.

Metadatamatic is available online: Metadatamatic

7.4.15 Evaluation of Explanations for Relational Graph Convolutional Network Link Prediction on Knowledge Graph

Participants:	Nicholas Halliwell, Fabien Gandon, Freddy Leccue, Serena Villata.

This collaboration is considering the need for empirical evaluation of explanation quality [66] and the fundamental problem of evaluating the quality of explanations generated by some of the latest AI systems[64].

We first proposed a rule-based and ontology based generator for a simplified benchmark focusing on providing non-ambiguous explanations for knowledge graph link prediction using Relational Graph Convolutional Networks (RGCN) [94, 84].

We then proposed an extended method to support user-scored evaluation of non-unique explanations for link prediction by RGCNs integrating the constraint of having multiple possible explanations for a prediction of different value for a user.

We also evaluated the impact of injecting ground truth explanations on relational graph convolutional networks and their explanation methods for link Prediction on knowledge Graphs [65].

This work on evaluating and improving explanation quality of graph neural network link prediction on knowledge graph was finally defended in the PhD of Nicholas Halliwell.[91]

7.4.16 Extending electronic medical records vector models with knowledge graphs to improve hospitalization prediction

Participants: Raphael Gazzoti, Catherine Faron, Fabien Gandon.

We proposed to address the problem of hospitalization prediction for patients with an approach that enriches vector representation of EMRs with information extracted from different knowledge graphs before learning and predicting. In addition, we performed an automatic selection of features resulting from knowledge graphs to distinguish noisy ones from those that can benefit the decision making. We evaluted our results with experiments on the PRIMEGE PACA database that contains more than 600,000 consultations carried out by 17 general practitioners (GPs). A statistical evaluation shows that our proposed approach improves hospitalization prediction. More precisely, injecting features extracted from cross-domain knowledge graphs in the vector representation of EMRs given as input to the prediction algorithm significantly increases the F1 score of the prediction. By injecting knowledge from recognized reference sources into the representation of EMRs, it is possible to significantly improve the prediction of medical events. [32]

8 Bilateral contracts and grants with industry

8.1 Bilateral contracts with industry

Curiosity Collaborative Project

Participants: Catherine Faron, Oscar Rodríguez Rocha, Molka Dhouib.

Partner: TeachOnMars.

This collaborative project with the TeachOnMars company started in October 2019. TeachOnMars is developping a platform for mobile learning. The aim of this project is to develop an approach for automatically indexing and semantically annotating heterogeneous pedagogical resources from different sources to build up a knowledge graph enabling to compute training paths, that correspond to the learner's needs and learning objectives.

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 Duput 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 will develop 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 will be carried out by Lucie Cadorel with the advice of Andrea Tettamanzi, is, 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.

9   Partnerships and cooperations

9.1   International initiatives

9.1.1   Associate Teams in the framework of an Inria International Lab or in the framework of an Inria International Program

PROTEMICS

> **Participants:**   Olivier Corby, Catherine Faron, Fabien Gandon, Iliana Petrova.

Title: Protégé and SHACL extension to support ontology validation

Duration: 2020 -> 2023

Coordinator: Rafael Gonçalves (rafael.goncalves@stanford.edu)

Partners:

- Stanford University Stanford (USA)

Inria contact: Fabien Gandon

Summary: We propose to investigate the extension of the structure-oriented SHACL validation to include more semantics, and to support ontology validation and the modularity and reusability of the associated constraints. Where classical Logical (OWL) schema validation focuses on checking the semantic coherence of the ontology, we propose to explore a language to capture ontology design patterns as extended SHACL shapes organized in modular libraries. The overall objective of our work is to augment the Protégé editor with fundamental querying and reasoning capabilities provided by CORESE, in order to assist ontology developers in performing ontology quality assurance throughout the life-cycle of their ontologies

9.2   International research visitors

9.2.1   Visits of international scientists

Other international visits to the team

Luciana Nedel

Status (PhD)

Institution of origin: University Federal do Rio Grande do Sul

Country: Brazil

Dates: September 2022

Context of the visit: collaboration on the topic of Virtual Reality

Mobility program/type of mobility: research stay, invited professor (1 month) funded by I3S lab

Paolo Buono

Status (PhD)

Institution of origin: University of Bari

Country: Italy

Dates: December 1st 2022

Context of the visit: invited lecture "What do two stars like L. Nimoy (Star Trek) and H. Ford (Star Wars) have in common? an answer provided by dynamic hypergraph visualization"

Mobility program/type of mobility: invited lecture

Félix Albertos Marco

Status (PhD)

Institution of origin: Universidad Castilla La Mancha

Country: Spain

Dates: May-July 2022

Context of the visit: collaboration on the topic of Web interaction

Mobility program/type of mobility: research stay research stay (3 month) funded by the Universidad Castilla La Mancha

Kevyn Collins-Thompson

Status (PhD)

Institution of origin: University of Michigan

Country: USA

Dates: June 13-16, 2022

Context of the visit: invited lecture "Search engines that help people learn"

Mobility program/type of mobility: invited lecture

9.3 European initiatives

AI4EU

Title: A European AI On Demand Platform and Ecosystem

Duration: *2019 - 2022*

Coordinator: THALES

Partners:

- AGENCIA ESTATAL CONSEJO SUPERIOR DEINVESTIGACIONES CIENTIFICAS (Spain)
- ALMA MATER STUDIORUM - UNIVERSITA DI BOLOGNA (Italy)
- ARISTOTELIO PANEPISTIMIO THESSALONIKIS (Greece)
- ASSOCIACAO DO INSTITUTO SUPERIOR TECNICO PARA A INVESTIGACAO E DESENVOLVIMENTO (Portugal)
- BARCELONA SUPERCOMPUTING CENTER - CENTRO NACIONAL DE SUPERCOMPUTACION (Spain)
- BLUMORPHO SAS (France)
- BUDAPESTI MUSZAKI ES GAZDASAGTUDOMANYI EGYETEM (Hungary)
- BUREAU DE RECHERCHES GEOLOGIQUES ET MINIERES (France)
- CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE CNRS (France)
- CINECA CONSORZIO INTERUNIVERSITARIO (Italy)
- COMMISSARIAT A L ENERGIE ATOMIQUE ET AUX ENERGIES ALTERNATIVES (France)
- CONSIGLIO NAZIONALE DELLE RICERCHE (Italy)
- DEUTSCHES FORSCHUNGSZENTRUM FUR KUNSTLICHE INTELLIGENZ GMBH (Germany)
- DEUTSCHES ZENTRUM FUR LUFT - UND RAUMFAHRT EV (Germany)
- EOTVOS LORAND TUDOMANYEGYETEM (Hungary)
- ETHNIKO KAI KAPODISTRIAKO PANEPISTIMIO ATHINON (Grecce)
- ETHNIKO KENTRO EREVNAS KAI TECHNOLOGIKIS ANAPTYXIS (Greece)
- EUROPEAN ORGANISATION FOR SECURITY (Belgium)
- FONDATION DE L'INSTITUT DE RECHERCHE IDIAP (Switzerland)
- FONDAZIONE BRUNO KESSLER (Italy)
- FORUM VIRIUM HELSINKI OY (Finland)
- FRANCE DIGITALE (France)

- FRAUNHOFER GESELLSCHAFT ZUR FOERDERUNG DER ANGEWANDTEN FORSCHUNG E.V. (Germany)
- FUNDACION CARTIF (Spain)
- FUNDINGBOX ACCELERATOR SP ZOO (Poland)
- FUNDINGBOX RESEARCH APS (Denmark)
- GOODAI RESEARCH SRO (Czech Republic)
- Hochschule für Technik und Wirtschaft Berlin (Germany)
- IDRYMA TECHNOLOGIAS KAI EREVNAS (Greece)
- IMT TRANSFERT (France)
- INSTITUT JOZEF STEFAN (Slovenia)
- INSTITUT POLYTECHNIQUE DE GRENOBLE (France)
- INTERNATIONAL DATA SPACES EV (Germany)
- KARLSRUHER INSTITUT FUER TECHNOLOGIE (Germany)
- KNOW-CENTER GMBH RESEARCH CENTER FOR DATA-DRIVEN BUSINESS & BIG DATA ANALYTICS (Austria)
- NATIONAL CENTER FOR SCIENTIFIC RESEARCH "DEMOKRITOS" (Greece)
- NATIONAL UNIVERSITY OF IRELAND GALWAY (Ireland)
- NORGES TEKNISK-NATURVITENSKAPELIGE UNIVERSITET NTNU (Norway)
- OFFICE NATIONAL D'ETUDES ET DE RECHERCHES AEROSPATIALES (France)
- ORANGE SA (France)
- OREBRO UNIVERSITY (Sweden)
- QWANT (France)
- TECHNICKA UNIVERZITA V KOSICIACH (Slovakia)
- TECHNISCHE UNIVERSITAET MUENCHEN (Germany)
- TECHNISCHE UNIVERSITAET WIEN (Austria)
- TECHNISCHE UNIVERSITAT BERLIN (Germany)
- THALES (France)
- THALES ALENIA SPACE FRANCE SAS (France)
- THALES SIX GTS FRANCE SAS (France)
- THOMSON LICENSING (France)
- TILDE SIA (Latvia)
- TWENTY COMMUNICATIONS SRO (Slovakia)
- UNIVERSIDAD POLITECNICA DE MADRID (Spain)
- UNIVERSIDADE DE COIMBRA (Portugal)
- UNIVERSITA CA' FOSCARI VENEZIA (Italy)
- UNIVERSITA DEGLI STUDI DI SIENA (Italy)
- UNIVERSITAT POLITECNICA DE CATALUNYA (Spain)
- UNIVERSITE DE LORRAINE (France)
- UNIVERSITE GRENOBLE ALPES (France)
- UNIVERSITY COLLEGE CORK - NATIONAL UNIVERSITY OF IRELAND, CORK (Ireland)
- UNIVERSITY OF LEEDS (UK)
- VRIJE UNIVERSITEIT BRUSSEL (Belgium)

- WAVESTONE (France)
- WAVESTONE ADVISORS (France)
- WAVESTONE LUXEMBOURG SA (Luxembourg)

Inria contact: Olivier Corby (for Wimmics)

Summary: In January 2019, the AI4EU consortium was established to build the first European Artificial Intelligence On-Demand Platform and Ecosystem with the support of the European Commission under the H2020 programme. The activities of the AI4EU project include:

- The creation and support of a large European ecosystem spanning the 28 countries to facilitate collaboration between all Europeans actors in AI (scientists, entrepreneurs, SMEs, Industries, funding organizations, citizens...);
- The design of a European AI on-Demand Platform to support this ecosystem and share AI resources produced in European projects, including high-level services, expertise in AI research and innovation, AI components and datasets, high-powered computing resources and access to seed funding for innovative projects using the platform;
- The implementation of industry-led pilots through the AI4EU platform, which demonstrates the capabilities of the platform to enable real applications and foster innovation;
- Research activities in five key interconnected AI scientific areas (Explainable AI, Physical AI, Verifiable AI, Collaborative AI, Integrative AI), which arise from the application of AI in real-world scenarios;
- The funding of SMEs and start-ups benefitting from AI resources available on the platform (cascade funding plan of €3m) to solve AI challenges and promote new solutions with AI;
- The creation of a European Ethical Observatory to ensure that European AI projects adhere to high ethical, legal, and socio-economical standards; click here to know more
- The production of a comprehensive Strategic Research Innovation Agenda for Europe
- The establishment of an AI4EU Foundation that will ensure a handover of the platform in a sustainable structure that supports the European AI community in the long run.

In the context of the AI4EU European project, we have translated the Thales Knowledge Graph into an RDF graph and have defined a set of SPARQL queries to query and navigate the graph. This has been integrated into the AI4EU endpoint prototype.

Web site: AI4EU Project

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 organisations) 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.

9.3.1   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

9.4   National initiatives

Ministry of Culture: MonaLIA 3.0

> **Participants:**   Anna Bobasheva, Fabien Gandon, Frédéric Precioso.

This work combines semantic reasoning and machine learning to create tools that allow curators of the visual art collections to identify and correct the annotations of the artwork as well as to improve the relevance of the content-based search results in these collections. The research is based on the Joconde database maintained by French Ministry of Culture that contains illustrated artwork records from main French public and private museums representing archeological objects, decorative arts, fine arts, historical and scientific documents, etc. The Joconde database includes semantic metadata that describes properties of the artworks and their content. The developed methods create a data pipeline that processes metadata, trains a Convolutional Neural Network image classification model, makes prediction for the entire collection and expands the metadata to be the base for the SPARQL search queries. We developed a set of such queries to identify noise and silence in the human annotations and to search image content with results ranked according to the relevance of the objects quantified by the prediction score provided by the deep learning model. We also developed methods to discover new contextual relationships between the concepts in the metadata by analyzing the contrast between the concepts similarities in the Joconde's semantic model and other vocabularies and we tried to improve the model prediction scores based on the semantic relations. Our results show that cross-fertilization between symbolic AI and machine learning can indeed provide the tools to address the challenges of the museum curators work describing the artwork pieces and searching for the relevant images [28].

ANR WASABI

> **Participants:**   Michel Buffa, Elena Cabrio, Catherine Faron, Alain Giboin.

The ANR project WASABI started in January 2017 with IRCAM, Deezer, Radio France and the SME Parisson, consists in building a 2 million songs knowledge base of commercial popular music (rock, pop, etc.). Its originality is the joint use of audio-based music information extraction algorithms, song lyrics analysis algorithms (natural language processing), and the use of the Semantic Web. Web Audio technologies will then explore these bases of musical knowledge by providing innovative applications for composers, musicologists, music schools, and sound engineers, music broadcasters and journalists. This project is in its mid-execution and gave birth to many publications in international conferences as well as some mainstream coverage (i.e for "la fête de la Science"). Participation in the ANR OpenMiage project aimed at offering online Bachelor and Master degrees.

The project also led to industrial transfer of some of the results (partnership with AmpedStudio.com/Amp Track company) for the integration of our software into theirs), SATT PACA.

Web site: Wasabi HomePage

ANR D2KAB

> **Participants:**   Olivier Corby, Catherine Faron, Fabien Gandon, Franck Michel, Na-
> dia Yacoubi Ayadi.

Partners: LIRMM, INRAE, IRD, ACTA

D2KAB is an ANR project which started in June 2019, 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 Sophia Antipolis-Méditerranée

DeKaloG (Decentralized Knowledge Graphs) 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, Xiaoou 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.

ANR CROQUIS

Participants:	Andrea Tettamanzi.

The ANR project CROQUIS started in March 2022 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.

Web site: CROQUIS Project

DBpedia.fr

Participants:	Fabien Gandon, Franck Michel, Célian Ringwald.

The DBpedia.fr project ensures the creation and maintenance of a French chapter of the DBpedia knowledge base. This project was the first project of the Semanticpedia convention signed by the Ministry of Culture, the Wikimedia foundation and Inria in 2012. Wimmics continued to maintain the Frecnh DBpedia chapter since.

A new project proposal was selected in 2021 between Inria and the Ministry of Culture to support evolutions and long-term sustaining of this project. The roadmap was spread over one year and planned to cover the following work packages:

Work package 1: Sustaining and increasing the visibility of the French-speaking chapter DBpedia; Automation of administration and maintenance tasks; Automation of the means of dissemination and adoption; Deployment in production of DBpedia Live and DBpédia Historic.

Work package 2: Evaluation of the cross-fertilisation of DBpedia with other Wikimedia Foundation projects (in particular Wikidata)

Work package 3: Development of new research areas including the application of NLP techniques to the text of Wikipedia articles and DBpedia Spotlight in French.

Convention between Inria and the Ministry of Culture

Participant: Fabien Gandon.

We supervise the research convention with the Ministry of Culture to foster research and development at the crossroad of culture and digital sciences. This convention signed between Inria and the Ministry of Culture provides a framework to support projects at the crossroad of the cultural domain and the digital sciences. The main visible event was the annual workshop "Atelier 2022 ministère de la Culture et Inria".

CovidOnTheWeb - Covid Inria program

Participants: Valentin Ah-Kane, Anna Bobasheva, Lucie Cadorel, Olivier Corby, Elena Cabrio, Jean-Marie Dormoy, Fabien Gandon, Raphaël Gazzotti, Alain Giboin, Santiago Marro, Tobias Mayer, Aline Menin, Franck Michel, Andrea Tettamanzi, Serena Villata, Marco Winckler.

The Covid-On-The-Web project aims to allow biomedical researchers to access, query and make sense of COVID-19 scholarly literature. To do so, we designed and implemented a pipeline levereding our skills in knowledge representation, text mining, argument mining and visualization techniques to process, analyze and enrich the COVID-19 Open Research Dataset (CORD-19) that gathers 100,000+ full-text scientific articles related to the coronaviruses.

The generated RDF dataset comprises the Linked Data description of (1) named entities (NE) mentioned in the CORD-19 corpus and linked to DBpedia, Wikidata and other BioPortal vocabularies, and (2) arguments extracted using ACTA, a tool automating the extraction and visualization of argumentative graphs, meant to help clinicians analyze clinical trials and make decisions.

On top of this dataset, we have adapted visualization and exploration tools (MGExplorer, Arviz) to provide Linked Data visualizations that meet the expectations of the biomedical community.
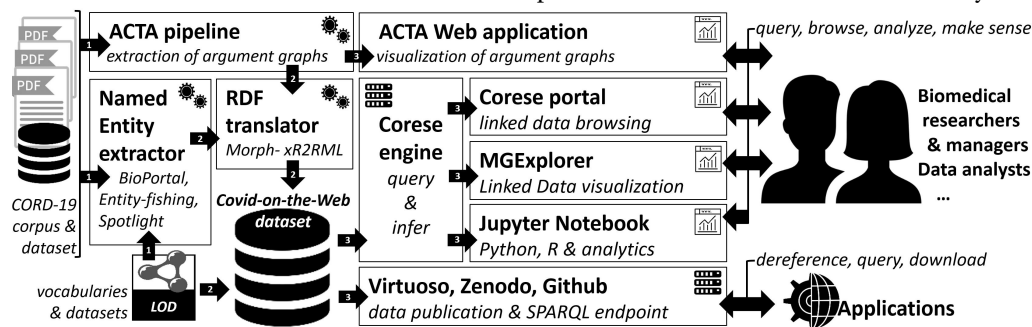
ISSA (AAP Collex-Persée)

Participants: Franck Michel, Marco Winckler, Anna Bobasheva, Olivier Corby.

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 primarily 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

9.5   Regional initiatives

3IA Côte d'Azur

Participants:	Elena Cabrio, Catherine Faron, Fabien Gandon, Freddy Limpens, Andrea Tettamanzi, Serena Villata.

3IA Côte d'Azur is one of the four "Interdisciplinary Institutes of Artificial Intelligence"[12] that were created in France in 2019. Its ambition is to create an innovative ecosystem that is influential at the local, national, and international level. The 3IA Côte d'Azur institute is led by Université Côte d'Azur in partnership with major higher education and research partners in the region of Nice and Sophia Antipolis: CNRS, Inria, INSERM, EURECOM, ParisTech MINES, and SKEMA Business School. The 3IA Côte d'Azur institute is also supported by ECA, Nice University Hospital Center (CHU Nice), CSTB, CNES, Data Science Tech Institute and INRAE. The project has also secured the support of more than 62 companies and start-ups.

We have four 3IA chairs for tenured researchers of Wimmics and several grants for PhD and postdocs.

We also have an industrial 3IA Affiliate Chair with the company Mnemotix focused on the industrialisation and scalability of the CORESE software.

UCA IDEX OTESIA project "Artificial Intelligence to prevent cyberviolence and hate speech online"

Participants:	Elena Cabrio, Serena Villata, Anais Ollagnier.

The project will cross the approaches of mediation / remediation: it is planned to develop a software to detect hate messages based on an analysis of natural language, but also to understand their argumentative structure (not a simple detection of isolated words, insults) and to develop the critical spirit of the victims and therefore, to do this, to develop a counter-speech. Hence interventions in 6 secondary schools for role-playing games that will serve as a basis for data collection. This data, once analysed, will support the development of software to detect hate and violent speech online. As part of a restitution of the work carried out, the institutions will participate in a collaborative manner in the development of counter-speech.

10   Dissemination

10.1   Promoting scientific activities

10.1.1   Scientific events: organisation

Participants:	Michel Buffa, Fabien Gandon, Marco Winckler.

---

[12]3IA Côte d'Azur Web site

General chair, scientific chair

- Marco Winckler (General Chair) of the 14th The ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS'2022), June 21-24, 2022, Sophia Antipolis, France.(programme)

Member of the organizing committees

- Marco Winckler (co-organizer) of the IFIP TC13 Open Symposium 2022 – York, UK – March 17th, 2022. (programme)

- Michel Buffa and Shihong Ren : organizers of the "WebAudio Plugin" workshop at the WebAudio Conference 2022 (WAC 2022), July 5-7, 2022.

- Fabien Gandon: co-organiser "Atelier ministère de la Culture et Inria", 22/03/2022 aux Archives Nationales, cf programme.

10.1.2 Scientific events: selection

Participants: Catherine Faron, Fabien Gandon.

Chair of conference program committees

- Catherine Faron: Program chair of EGC 2023 (conférence francophone sur l'Extraction et la Gestion des Connaissances), January 16-20, 2023, programme

Member of the conference program committees

- Marco Winckler: Advanced Visual Interfaces (AVI 2022), ManComp 2022, EUROVIS 2022, HCSE 2022 (International Working Conference on Human-Centered Software Engineering), ACM EICS (ACM SIGCHI Symposium on Engineering Interactive Computing Systems - Associated Chair), IEEE ETFA 2022 (The 27th Annual Conference of the IEEE Industrial Electronics Society), ICWE 2022 (International Conference on Web Engineering, IHC 2022 (Brazilian Symposium on Human-Computer), Interacción 2022, IFIP IOT 2022, ISD 2022 (International Conference on Information System Development), ACM IUI (International Conference on Intelligent User Interfaces), NordiCHI 2022, S-BPM ONE, WISE 2022 (International Conference of Web Information Systems Engineering (WISE 2022).

- Catherine Faron : PC member of ESWC 2022 (European Semantic Web Conference), EKAW 2022 (International Conference on Knowledge Engineering and Knowledge Management), ISWC 2022 (International Semantic Web Conference), TheWebConf 2022 (The Web Conference), SEMANTiCS 2022 (International Conference on Semantics Systems), ICCS 2022 (International Conference on Conceptual Structures), IC 2022 (Journées francophones d'Ingénierie des Connaissances)

- Fabien Gandon : PC member of AAAI 2022 (Association for the Advancement of Artificial Intelligence), ESWC 2022 (European Semantic Web Conference), EKAW 2022 (International Conference on Knowledge Engineering and Knowledge Management), IJCAI-ECAI 2022 (International Joint Conference on Artificial Intelligence), ISWC 2022 (International Semantic Web Conference) ....

- Andrea Tettamanzi: PC member of TheWebConf 2022, ICAART 2022, EGC 2023, WI-IAT 2022, SUM 2022, EKAW 2022, UAI 2022,

- Serena Villata: Senior Program Committee member for the AAMAS-2022 Innovative Applications Area. Since September 2021, Action Editor for the ACL Rolling Review.

- Elena Cabrio : Member of the Senior Program Committee of ACL ARR Rolling Reviews 22, IJCAI-ECAI 22, IJCAI 2022.

10.1.3 Journal

Participants: Catherine Faron, Fabien Gandon, Marco Winckler, Elena Cabrio.

Member of the editorial boards

- Catherine Faron: board member of Revue Ouverte d'Intelligence Artificielle (ROIA)

- Elena Cabrio: Member of the Editorial Board of the Italian Journal of Computational Linguistics (IJCoL ISSN 2499-4553).

- Marco Winckler: Member of the Editorial Board of Journal of Web Engineering, Interacting with Computers (Oxford Press - ISSN 1873-7951), Multimodal Technologies and Interaction (ISSN 2414-4088), Behaviour & Information Technology, PACM Proceedings on Human-Computer Interaction (ACM Sheridam), IFIP « Advances in Information and Communication Technology » (Springer).

Reviewer - reviewing activities

- Catherine Faron: Semantic Web journal, Journal of Web Semantics

10.1.4 Invited talks

Participants: Fabien Gandon, Serena Villata, Elena Cabrio.

- Fabien Gandon:

  - EKAW 2022 Keynote on "A shift in our research focus: from knowledge acquisition to knowledge augmentation", 28/09/2022, see recording.

- Serena Villata:

  - Keynote speaker at the the 24th International Conference on Principles and Practice of Multi-Agent Systems (PRIMA-2022), November 2022, Valencia, Spain.[13]

  - Keynote speaker at the 20th International Workshop on Non-Monotonic Reasoning (NMR-2022) on "Fallacious arguments: the Place Where Knowledge Representation and Argument Mining Meet Each Other", August 2022, Haifa, Israel.[14]

  - Keynote speaker at the 2nd International Workshop on Knowledge Graphs for Online Discourse Analysis (BeyondFacts-2022) on "Towards argument-based explanatory dialogues: from argument mining to argument generation", April 2022, online event.[15]

- Elena Cabrio

  - Invited speaker: "Processing Natural Language to Extract, Analyze and Generate Knowledge and Arguments from Texts". Seminar. ALMAnaCH research team, INRIA Paris. March 2022.

---

[13]PRIMA-2022
[14]NMR-2022
[15]BeyondFacts-2022

10.1.5 Leadership within the scientific community

Participants: Catherine Faron, Fabien Gandon, Marco Winckler, Serena Villata.

- Fabien Gandon was a member, until November 2022, of the Semantic Web Science Association (SWSA) a non-profit organisation for promotion and exchange of the scholarly work in Semantic Web and related fields throughout the world ans steering committee of the ISWC conference.

- Catherine Faron is a member of the steering committee of the EKAW conference

- Marco Winckler is French representative at the IFIP TC13 and member of the Association Francophone pour l'Interaction Homme-Machine.

- Serena Villata is a member of the Comité Nationale Pilote d'Ethique du Numerique (CNPEN) since December 2019.

10.1.6 Scientific expertise

Participants: Catherine Faron, Fabien Gandon, Serena Villata.

- Catherine Faron was an expert for the MSCA Postdoctoral Fellowships call.

- Fabien Gandon is a member of the Choose France committee, was an experts for the Canada Excellence Research Chairs (CERC) competition 2022 and an expert for a call of the European Science Foundation, member of the Polifonia H2020 project advisory board.

- Serena Villata, from November 2021 to November 2022, has been mandated (chargée de mission) for the Culture Ministry for the mission on "Conversational Agents", together with Célia Zolynski and Karine Favro.[16]

- Serena Villata is a member of the scientific committee CE-23 of the Agence Nationale de la Recherche (ANR) for the AAPG 2022 (7 projects to review, participation to the jury meetings).

10.1.7 Research administration

Participants: Elena Cabrio, Olivier Corby, Catherine Faron, Fabien Gandon, Andrea Tettamanzi, Serena Villata, Marco Winckler.

- Olivier Corby: member of the working group on environmental issue at I3S laboratory.

- Catherine Faron: member of the I3S laboratory council; member of the steering committee of the AFIA college on Knowledge Engineering.

- Fabien Gandon : Leader of the Wimmics team ; Vice-director of Research Inria Sophia Antipolis until September ; co-president of scientific and pedagogical council of the Data Science Technical Institure (DSTI) ; Member of the Evaluation Committee of Inria until September; W3C Advisory Committee Representative (AC Rep) for Inria ; Leader of the convention Inria - Ministry of Culture.

- Serena Villata: Deputy Scientific Director of 3IA Côte d'Azur Institute.

- Elena Cabrio is member of the Conseil d'Administration (CA) of the French Association of Computational Linguistics (ATALA) ; Member of the Bureau of the Académie 1 of IDEX UCA JEDI.

---

[16]Culture Ministry Webpage about the mission.

- Andrea Tettamanzi is the leader of the SPARKS team at I3S laboratory.

- Marco Winckler is joint-director of the SPARKS team at the I3S Laboratory, Secretary for the IFIP TC13 on Human-Computer Interaction, member of the Steering Committee of INTERACT.

10.2   Teaching - Supervision - Juries

10.2.1   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.

- 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 UCA 25 hours Semantic Web

  – Licence 3 IA DS4H UCA 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 UCA: 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 DSTI: Data pipeline, 50h.

- Fabien Gandon :

  – Master: Integrating Semantic Web technologies in Data Science developments, 78 h, M2, DSTI, France.

  – Tutorial Soph.IA Master Classes: "INRIA : CORESE, la plate-forme logicielle Open Source pour des graphes de connaissance aux standards du Web", 2h, see (online resources).

- Aline Menin :

  – Master 1, Data Sciences & Artificial Intelligence, UCA, 14h (CM/TP), Data visualization.

  – Polytech 5ème année, UCA, 13.5h (CM/TP), Data visualization.

  – BUT 2, IUT Nice Côte d'Azur, 106h éq. TD, Dévéloppement efficace & Qualité de développement.

- Amaya Nogales Gómez:

  – Master 1, Data Sciences & Artificial Intelligence, UCA, 20h (CM/TD), Security and Ethical Aspects of Data.

  – Licence 2, Licence Informatique, UCA, 36h (TP), Structures de données et programmation C.

- 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.

  - Master Communication et Langage Politique - University Côte d'Azur: Argumentation, 15 hours (CM+TD), 10 students.

- Elena Cabrio:

  - Master I Computer Science, Text Processing in AI, e. 30 hours (eq. TD).

  - Master 2 MIAGE IA - University Côte d'Azur: I.A. et Langage : Traitement automatique du langage naturel, 28 hours (CM+TP), 30 students.

  - 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.

  - Master Modeling for Neuronal and Cognitive Systems. Text analysis, deep learning and statistics, 18.5 hours.

  - License 2, Computer Science. Web Technologies, 54 hours.

- Andrea Tettamanzi

  - Licence: Introduction à l'Intelligence Artificielle, 45 h ETD, L2, UCA, France.

  - Master: Logic for AI, 30 h ETD, M1, UCA, France.

  - Master: Web, 30 h ETD, M1, UCA, France.

  - Master: Algorithmes Évolutionnaires, 24.5 h ETD, M2, UCA, France.

  - Master: Modélisation del l'Incertitude, 24.5 h ETD, M2, UCA, France.

- Marco Winckler

  - Licence 3: Introduction to Human-Computer Interaction, 45 h ETD, UCA, Polytech Nice, France.

  - Master 1: Accessibility and Universal Design, 10 h ETD, UCA, DS4H, France.

  - Master 1: Methods and tools for technical and scientific writing, Master DSAI, 15 h ETD, UCA, DS4H, France.

  - Master 1: Introduction to Information Visualization, Master DSAI, 15 h ETD, UCA, DS4H, France.

  - Master 2: Introduction to Scientific Research, 10 h ETD, UCA, DS4H, France.

  - Master 2: Introduction to Scientific Research, 15 h ETD, UCA, Polytech Nice, France.

  - Master 2: Data Mining Visualisation, 8 h ETD, UCA, Polytech Nice, France.

  - Master 2: Data Visualization, 15 h ETD, UCA, MBDS DS4H, France.

  - Master 2: Design and Evaluation of User Interfaces, 45 ETD, UCA, Polytech Nice, France.

  - Master 2: Multimodal Interaction Techniques, 15 ETD, UCA, Polytech Nice, France.

  - Master 2: coordination of the TER (Travaux de Fin d'Etude), UCA, Polytech Nice, France.

  - Master 2: coordination of the track on Human-Computer Interaction at the Informatics Department, UCA, Polytech Nice, France.

- Molka Dhouib

– Licence 3/3A SI PNS: Relational Databases, 32h (TD).

- Molinet Benjamin:

  – License 2, Computer Science. Web Technologies, 54 hours.

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 (2015-2022), 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 (2015-2022), 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 (2015-2022), MOOC page.

10.2.2 Supervision

Participants: Michel Buffa, Elena Cabrio, Catherine Faron, Fabien Gandon, Aline Menin, Andrea Tettamanzi, Serena Villata, Marco Winckler, Nadia Yacoubi Ayadi.

- PhD in progress: Ali Ballout, *Active Learning for Axiom Discovery*, Andrea Tettamanzi, UCA.

- PhD in progress: 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*, Andrea Tettamanzi, UCA.

- PhD defended: *Ahmed El Amine Djebri, Uncertainty in Linked Data*, UCA, Andrea Tettamanzi, Fabien Gandon [90].

- PhD in progress: Rony Dupuy Charles, *Combinaison d'approches symboliques et connexionnistes d'apprentissage automatique pour les nouvelles méthodes de recherche et développement en agro-végétale-environnement*, Andrea Tettamanzi, UCA.

- PhD defended: Antonia Ettore, *Artificial Intelligence for Education and Training: Knowledge Representation and Reasoning for the development of intelligent services in pedagogical environment*, UCA, Catherine Faron, Franck Michel.

- PhD in progress: Rémi Felin, *Découverte évolutive d'axiomes à partir de graphes de connaissances*, UCA, Andrea Tettamanzi, Catherine Faron.

- PhD defended: Nicholas Halliwell, *Explainable and Interpretable Prediction*, UCA, Fabien Gandon [91].

- PhD in progress: Santiago Marro, *Argument-based Explanatory Dialogues for Medicine*, UCA 3IA, Elena Cabrio and Serena Villata.

- PhD in progress: Benjamin Molinet, *Explanatory argument generation for healthcare applications*, UCA 3IA, Elena Cabrio and Serena Villata.

- PhD in progress: Pierpaolo Goffredo, *Fallacious Argumentation in Political Debates*, UCA 3IA, Elena Cabrio and Serena Villata.

- PhD in progress: Xiaoou Wang, *Counter-argumentation generation to fight online disinformation*, UCA, Elena Cabrio and Serena Villata.

- PhD in progress: *Benjamin Ocampo, Subtle and Implicit Hate Speech Detection*, UCA 3IA, Elena Cabrio and Serena Villata.

- PhD in progress: Maroua Tikat, *Visualisation multimédia interactive pour l'exploration d'une base de métadonnées multidimensionnelle de musiques populaires.* Michel Buffa, Marco Winckler.

- PhD in progress: Célian Ringwald, *Learning RDF pattern extractors for a language from dual bases Wikipedia/LOD*, Fabien Gandon, Franck Michel, 3IA,UCA.

- PhD in progress: Florent Robert, *Analyzing and Understanding Embodied Interactions in Extended Reality Systems.* Co-supervision with Hui-Yin Wu, Lucile Sassatelli, and Marco Winckler.

- PhD in progress: Clément Quere. *Immersive Visualization Techniques for spatial-temporal data.* Co-supervision with Aline Menin and Hui-Yin Wu.

Internships

- Master 2 internship: Arnaud Barbe . Co-supervised by Catherine Faron, Franck Michel, Nadia Yacoubi Ayadi.

- Master 2 internship: Felipe Gomes "Développement de la plateforme ANR CROBORA (CROssing BORders Archives)", Internship at Université Côte d'Azur – Laboratoire de recherche SIC. Lab Méditerranée, France. Co-supervision with Matteo Treleani.

Master Projects (TER)

- Master 2 TER: ELMehdi Soummer. Co-supervised by Molka Dhouib, Catherine Faron.

- Master 1 TER: Ekaterina Kostrykina. Co-supervised by Molka Dhouib, Catherine Faron.

10.2.3 Juries

Participants: Catherine Faron, Serena Villata, Serena Villata, Marco Winckler.

- Catherine Faron was the president of a selection committee for the recruitment of a senior lecturer (maître de conférences) at Université Côte d'Azur, member of the Inria Comité Nice for the recruitment of post-doctoral researchers and delegations, member of the PhD defense juries of Thamer Mecharnia (reviewer) and Nicolas Lasolle (reviewer).

- Serena Villata was a member of the PhD jury of Jorge Fernandez Davila, University of Toulouse. Title of the thesis: "Logic-based cognitive planning: from theory to implementation".

- Elena Cabrio was a member of the PhD jury of I. Sucameli, University of Pisa (Italy); K. Zarei, Institut Polytechnique de Paris (IP-Paris, France); S. Frenda, University of Turin (Italy); F. Ruggeri, University of Bologna (Italy).

- Marco Winckler was a member of the PhD Jury of Haritz MEDINA CAMACHO, Haritz. Title of the thesis: "Harnessing customization in Web Annotation: A Software Product Line approach". Presented at the Universidad del Pais Vasco (UPV/EHU), San Sebastien, Spain, on October 7th 2022. (Rapporteur).

10.3 Popularization

Participants: Fabien Gandon.

10.3.1 Articles and contents

Fabien Gandon:

- co-authored a paper on the history and evolution of the Web titled "A Never-Ending Project for Humanity Called "the Web"" [59].

- authored a novel to introduce recursivity in the SIF review: "Cours toujours" [98].

- IA : bâtir la confiance et garantir la souveraineté, INRIA interview

10.3.2 Interventions

- Fabien Gandon: performed 3 sessions "Chiche!" at Lycée Thierry Maulnier, 13/12/2022

11 Scientific production

11.1 Major publications

[1] D. Allemang, J. Hendler and F. Gandon. *Semantic Web for the Working Ontologist.* 3. ACM, June 2020. DOI: 10.1145/3382097. URL: https://hal.inria.fr/hal-02939606.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] N. Marie. 'Linked data based exploratory search'. Université Nice Sophia Antipolis, 12th Dec. 2014. URL: https://tel.archives-ouvertes.fr/tel-01130622.

[17] T. Mayer. 'Argument Mining on Clinical Trials'. Université Côte d'Azur, 16th Dec. 2020. URL: https://tel.archives-ouvertes.fr/tel-03209489.

[18] 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.

[19] 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.

[20] 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. URL: https://hal.archives-ouvertes.fr/hal-02939363.

[21] 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.

[22] 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. URL: https://hal.inria.fr/hal-02936130.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

11.2   Publications of the year

International journals

[28]   A. Bobasheva, F. Gandon and F. Precioso. 'Learning and Reasoning for Cultural Metadata Quality'.
         In: *Journal on Computing and Cultural Heritage* 15.3 (Sept. 2022). DOI: 10.1145/3485844. URL:
         https://hal.science/hal-03363442.

[29]   P. A. Bonatti, I. M. Petrova and L. Sauro. 'Optimizing the computation of overriding in DLN'. In:
         Artificial Intelligence 311 (Oct. 2022), p. 103764. DOI: 10.1016/j.artint.2022.103764. URL:
         https://hal.science/hal-03934420.

[30]   G. Bosetti, A. Tacuri, I. Gambo, S. Firmenich, G. Rossi, M. Winckler and A. Fernandez. 'ANDES: An
         approach to embed search services on the Web browser'. In: *Computer Standards and Interfaces*
         82 (Aug. 2022), p. 103633. DOI: 10.1016/j.csi.2022.103633. URL: https://hal.inria.fr/h
         al-03811257.

[31]   M. Fell, E. Cabrio, M. Tikat, F. Michel, M. Buffa and F. Gandon. 'The WASABI Song Corpus and
         Knowledge Graph for Music Lyrics Analysis'. In: *Language Resources and Evaluation* (2022). URL:
         https://hal.science/hal-03812106.

[32]   R. Gazzotti, C. Faron Zucker, F. Gandon, V. Lacroix-Hugues and D. Darmon. 'Extending electronic
         medical records vector models with knowledge graphs to improve hospitalization prediction'. In:
         Journal of Biomedical Semantics 13.6 (22nd Feb. 2022), p. 20. DOI: 10.1186/s13326-022-00261-
         9. URL: https://hal.science/hal-03502001.

[33]   C. Leturc and G. Bonnet. 'Reasoning about manipulation in multi-agent systems'. In: *Journal of
         Applied Non-Classical Logics* (30th Sept. 2022), pp. 1–67. DOI: 10.1080/11663081.2022.212406
         7. URL: https://hal.science/hal-03793389.

[34]   M. Louarn, F. Chatonnet, X. Garnier, T. Fest, A. Siegel, C. Faron and O. Dameron. 'Improving
         reusability along the data life cycle: a Regulatory Circuits Case Study'. In: *Journal of Biomedical
         Semantics* 13.1 (2022), pp. 1–11. DOI: 10.1186/s13326-022-00266-4. URL: https://hal.inri
         a.fr/hal-03602177.

[35]   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 Suits'. In: *Journal of Web Semantics*
         (19th Jan. 2023). DOI: 10.1016/j.websem.2023.100775. URL: https://hal.science/hal-03
         946680.

[36]   A. Menin, M. N. Do, C. Dal Sasso Freitas, O. Corby, C. Faron Zucker, A. Giboin and M. Winckler.
         'Using Chained Views and Follow-up Queries to Assist the Visual Exploration of the Web of Big
         Linked Data'. In: *International Journal of Human-Computer Interaction* (23rd Aug. 2022). DOI:
         10.1080/10447318.2022.2112529. URL: https://hal.science/hal-03518845.

[37]   S. Ren, L. Pottier, M. Buffa and Y. Yu. 'JSPatcher, a Visual Programming Environment for Building
         High-Performance Web Audio Applications'. In: *Journal of the Audio Engineering Society* (16th Nov.
         2022). URL: https://hal.inria.fr/hal-03871500.

[38]   S. Sarkadi, A. G. B. Tettamanzi and F. Gandon. 'Interoperable AI: Evolutionary Race Towards
         Sustainable Knowledge Sharing'. In: *IEEE Internet Computing* (28th Oct. 2022). DOI: 10.1109
         /MIC.2022.3214378. URL: https://hal.inria.fr/hal-03807461.

[39]   S. Villata, M. Araszkiewicz, K. Ashley, T. Bench-Capon, L. K. Branting, J. Conrad and A. Wyner.
         'Thirty years of artificial intelligence and law: the third decade'. In: *Artificial Intelligence and Law*
         30.4 (Dec. 2022), pp. 4143–4149. DOI: 10.1007/s10506-022-09327-6. URL: https://hal.sci
         ence/hal-03873418.

[40]   M. Winckler, P. Palanque, J.-L. Hak, E. Barboni, O. Nicolas and L. Goncalves. 'Engineering Anno-
         tations: A Generic Framework For Gluing Design Artefacts in Models of Interactive Systems'. In:
         Proceedings of the ACM on Human-Computer Interaction. Journal PACM EICS series EICS 2022
         (June 2022). URL: https://hal.inria.fr/hal-03657095.

[41]   H.-Y. Wu, F. A. S. Robert, T. Fafet, B. Graulier, B. Passin-Cauneau, L. Sassatelli and M. Winckler. 'Designing Guided User Tasks in VR Embodied Experiences'. In: *Proceedings of the ACM on Human-Computer Interaction* 6.158 (2022), pp. 1–24. DOI: 10.1145/3532208. URL: https://hal.inria.fr/hal-03635452.

International peer-reviewed conferences

[42]   A. Ballout, C. da Costa Pereira and A. G. B. Tettamanzi. 'Learning to Classify Logical Formulas Based on Their Semantic Similarity'. In: *Lecture Notes in Computer Science. Lecture Notes in Artificial Intelligence*. PRIMA 2022 - Principles and Practice of Multi-Agent Systems - 24th International Conference. Vol. LNCS. LNAI-13753. PRIMA 2022: Principles and Practice of Multi-Agent Systems. Valencia, Spain: Springer International Publishing, 12th Nov. 2023, pp. 364–380. DOI: 10.1007/978-3-031-21203-1_22. URL: https://hal.inria.fr/hal-03913481.

[43]   A. Ballout, A. G. B. Tettamanzi and C. da Costa Pereira. 'Predicting the Score of Atomic Candidate OWL Class Axioms'. In: WI-IAT 2022 - 21st IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. Niagara Falls, Canada, 17th Nov. 2022. URL: https://hal.science/hal-03905409.

[44]   D. Batisse, A. Vidal-Mazui, J.-P. Carlens, A. Cousson and M. Buffa. 'Host and Plugins : Parameter Automation Without Crossing the Audio Thread Barrier'. In: WAC 2022 - Web Audio Conference 2022. Cannes, France, 6th July 2022. URL: https://hal.inria.fr/hal-03812916.

[45]   J. Bruyat, P.-A. Champin, L. Médini and F. Laforest. 'Shacled Turtle: Schema-Based Turtle Auto-Completion'. In: Workshop on Visualization and Interaction for Ontologies and Linked Data 2022, co-located with the International Semantic Web Conference 2022. Vol. 3253. CEUR Workshop Proceedings. Hangzhou (Virtual Event), China, 28th Oct. 2022, pp. 2–15. URL: https://hal.science/hal-03834320.

[46]   M. Buffa. 'Web Audio Modules 2.0: Audio Plugins for the Web Platform!' In: ADC 2022 - Audio Developer Conference. Londres, United Kingdom, 14th Nov. 2022. URL: https://hal.inria.fr/hal-03871660.

[47]   M. Buffa, P. Kouyoumdjian, Q. Beauchet, Y. Forner and M. Marynowic. 'Making a guitar rack plugin -WebAudio Modules 2.0'. In: Web Audio Conference 2022. Cannes, France, 6th July 2022. URL: https://hal.inria.fr/hal-03812948.

[48]   M. Buffa, L. Pottier, S. Ren and Y. Yu. 'Making Computer Music on the Web with JSPatcher'. In: SMC 2022 - Sound and Music Computing 2022. Saint-Etienne, France, 5th June 2022. URL: https://hal.inria.fr/hal-03812972.

[49]   M. Buffa, S. Ren, O. Campbell, J. Kleimola, O. Larkin and T. Burns. 'Web Audio Modules 2.0: An Open Web Audio Plugin Standard'. In: *Proceedings of the WWW '22: The ACM Web Conference 2022*. WWW '22: The ACM Web Conference 2022. Virtual Event, France: ACM, 25th Apr. 2022, pp. 364–369. DOI: 10.1145/3487553.3524225. URL: https://hal.inria.fr/hal-03951799.

[50]   L. Cadorel, A. Blanchi and A. G. B. Tettamanzi. 'Connaissances géospatiales dans les annonces immobilières : détection et extraction d'information spatiale à partir du texte'. In: IC 2022 - Journées francophones d'Ingénierie des Connaissances (dans le cadre de PFIA 2022). IC, Journées francophones d'Ingénierie des Connaissances, PFIA 2022. Saint-Étienne, France: AfIA, 29th June 2022, pp. 20–21. URL: https://hal.inria.fr/hal-03927871.

[51]   L. Cadorel, D. Overal and A. G. B. Tettamanzi. 'Fuzzy representation of vague spatial descriptions in real estate advertisements'. In: Proceedings of the 6th ACM SIGSPATIAL International Workshop on Location-based Recommendations, Geosocial Networks and Geoadvertising, LocalRec 2022. Seattle Washington, United States: ACM, Nov. 2022, pp. 1–4. DOI: 10.1145/3557992.3565994. URL: https://hal.inria.fr/hal-03913497.

[52]   L. Cadorel, A. G. B. Tettamanzi and F. Gandon. 'Towards a representation of uncertain geospatial information in knowledge graphs'. In: GKG 2022 - Proceedings of the 1th ACM SIGSPATIAL International Workshop on Geospatial Knowledge Graphs. Seattle, Washington, United States, 1st Nov. 2022, pp. 1–2. DOI: 10.1145/3557990.3567588. URL: https://hal.inria.fr/hal-03913504.

[53] C. da Costa Pereira, D. Dubois, H. Prade and A. G. B. Tettamanzi. 'Parsimonious Representation of Knowledge Uncertainty using Metadata about Validity and Completeness'. In: *Proceedings of the 14th International Conference on Agents and Artificial Intelligence, ICAART 2022, Volume 2, Online Streaming, February 3–5, 2022*. ICAART 2022 - 14th International Conference on Agents and Artificial Intelligence. Vol. 2. Proceedings of the 14th International Conference on Agents and Artificial Intelligence - (Volume 2). Online Streaming, Portugal: SCITEPRESS : Science and Technology Publications, 3rd Feb. 2022, pp. 441–449. DOI: 10.5220/0010874000003116. URL: https://hal.science/hal-03719603.

[54] S. d'Angelo, F. Precioso and F. Gandon. 'Revisiting Artistic Style Transfer for Data Augmentation in A Real-Case Scenario'. In: IEEE ICIP 2022 - 29th IEEE International Conference on Image Processing. Bordeaux, France: IEEE, 16th Oct. 2022, pp. 4178–4182. DOI: 10.1109/ICIP46576.2022.9897728. URL: https://hal.science/hal-03921565.

[55] M. Di Russo, Z. Babutsidze, C. da Costa Pereira, M. Iacopetta and A. G. B. Tettamanzi. 'Agent-Based Modeling for Studying the Spontaneous Emergence of Money'. In: WI-IAT '22: The 21st IEEE/WIC/ACM International Conference on Web Intelligence and Agent Technologies. Niagara Falls, Canada: ACM, 17th Nov. 2022. URL: https://hal.inria.fr/hal-03913561.

[56] A. Ettorre, A. Bobasheva, F. Michel and C. Faron. 'Stunning Doodle: a Tool for Joint Visualization and Analysis of Knowledge Graphs and Graph Embeddings'. In: ESWC 2022 - European Semantic Web Conferences. Hersonissos, Greece, 29th May 2022. URL: https://hal.science/hal-03601145.

[57] A. Ettorre, A. Bobasheva, F. Michel and C. Faron. 'Stunning Doodle: un outil pour la visualisation et l'analyse conjointe de graphes de connaissances et leurs plongements'. In: PFIA 2022 - Journées francophones d'Ingénierie des Connaissances. Saint-Étienne, France, 2022. URL: https://hal.science/hal-03889910.

[58] R. Felin, O. Corby, C. Faron and A. G. B. Tettamanzi. 'Optimizing the Computation of a Possibilistic Heuristic to Test OWL SubClassOf Axioms Against RDF Data'. In: The 21st IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. Niagara Falls, France, 17th Nov. 2022. URL: https://hal.inria.fr/hal-03892080.

[59] F. Gandon and W. Hall. 'A Never-Ending Project for Humanity Called "the Web"'. In: WWW 2022 - ACM Web Conference. Lyon (virtual), France, 25th Apr. 2022. DOI: 10.1145/3485447.3514195. URL: https://hal.inria.fr/hal-03633526.

[60] 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. URL: https://hal.science/hal-03873412.

[61] D. Graux and A. Ollagnier. 'LOV-ES: Guiding the Ontology Selection to Structure Textual Data using Topic Modeling'. In: International Semantic Web Conference. Hangzou, China, 23rd Oct. 2022. URL: https://hal.science/hal-03765857.

[62] Q. Guimard, F. Robert, C. Bauce, A. Ducreux, L. Sassatelli, H.-Y. Wu, M. Winckler and A. Gros. 'On the link between emotion, attention and content in virtual immersive environments'. In: *2022 IEEE International Conference on Image Processing (ICIP) - IEEEXplore*. ICIP 2022 - IEEE International Conference on Image Processing. Bordeaux, France, 16th Oct. 2022. DOI: 10.1109/ICIP46576.2022.9897903. URL: https://hal.science/hal-03825059.

[63] Q. Guimard, F. Robert, C. Bauce, A. Ducreux, L. Sassatelli, H.-Y. Wu, M. Winckler and A. Gros. 'PEM360: A dataset of 360° videos with continuous Physiological measurements, subjective Emotional ratings and Motion traces'. In: MMSys 2022 - 13th ACM Multimedia Systems Conference. Athlone, Ireland, 14th June 2022. DOI: 10.1145/3524273.3532895. URL: https://hal.science/hal-03710323.

[64] N. Halliwell. 'Evaluating Explanations of Relational Graph Convolutional Network Link Predictions on Knowledge Graphs'. In: AAAI 2022 - 36th AAAI Conference on Artificial Intelligence. Vancouver, Canada, 22nd Feb. 2022. URL: https://hal.science/hal-03454121.

[65] N. Halliwell, F. Gandon and F. Lecue. 'Impact of Injecting Ground Truth Explanations on Relational Graph Convolutional Networks and their Explanation Methods for Link Prediction on Knowledge Graphs'. In: WI-IAT 2022 - The 21st IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. Niagara Falls / Hybrid, Canada, 17th Nov. 2022. URL: https://hal.science/hal-03771424.

[66] N. Halliwell, F. Gandon, F. Lecue and S. Villata. 'The Need for Empirical Evaluation of Explanation Quality'. In: AAAI 2022 - Workshop on Explainable Agency in Artificial Intelligence. Vancouver, Canada, 22nd Feb. 2022. URL: https://hal.science/hal-03591012.

[67] J. Lebrun, M. Buffa and J. Sintes. 'From blues to rock to jazz: three different web audio tube guitar amplifier simulator plugins'. In: *Proceedings of 7th Web Audio Conference (WAC 2022), Cannes, France*. WAC 2022 - Web Audio Conference 2022. Cannes, France, 6th July 2022. DOI: 10.5281/ZENODO.6769408. URL: https://hal.science/hal-03726153.

[68] A. Locoro, P. Buono, M. Winckler and S. Corchs. 'Representable AI: Towards a Unified View of Core Dimensions for a Visual Framework'. In: *CEUR proceedings 3136: https://dblp.org/db/conf/avi/copda2022.html*. Workshop on Cultures of Participation in the Digital Age: AI for Humans or Humans for AI ? Frascati (Rome), Italy, 7th June 2022. URL: https://hal.inria.fr/hal-03936512.

[69] S. Marro, E. Cabrio and S. Villata. 'Argumentation Quality Assessment: an Argument Mining Approach'. In: Argumentation Quality Assessment: an Argument Mining Approach. Rome, Italy, 28th Oct. 2022. URL: https://hal.science/hal-03934466.

[70] S. Marro, E. Cabrio and S. Villata. 'Graph Embeddings for Argumentation Quality Assessment'. In: *Graph Embeddings for Argumentation Quality Assessment*. EMNLP 2022 - Conference on Empirical Methods in Natural Language Processing. Abu Dhabi, United Arab Emirates, 7th Dec. 2022. URL: https://hal.science/hal-03934474.

[71] C. Martinie, P. Palanque and M. Winckler. 'Usability and User Experience are not Enough: Gaps to Fill to Design for and Assess Well-Being and Engagement'. In: *à paraître*. VIII Iberoamerican Conference on Human-Computer Interaction 2022. La Havane, Cuba, 13th Oct. 2022, pp. 1–14. URL: https://hal.inria.fr/hal-03811274.

[72] A. Menin, M. Buffa, M. Tikat, B. Molinet, G. Pelerin, L. Pottier, F. Michel and M. Winckler. 'Incremental and multimodal visualization of discographies: exploring the WASABI music knowledge base'. In: WAC 2022 - Web Audio Conference 2022. Vol. 2022. WAC Proceedings 2022. Cannes, France, 6th June 2022. DOI: 10.5281/zenodo.6767530. URL: https://hal.science/hal-03748134.

[73] F. Michel, F. Amardeilh, R. Bossy, C. Faron, C. Roussey and C. Noûs. 'Alignement entre sources : cas d'usage des plantes cultivées'. In: PFIA 2022 - Journées francophones d'Ingénierie des Connaissances. Saint-Étienne, France, 2022. URL: https://hal.science/hal-03889780.

[74] F. Michel, A. Toulet, A. Bobasheva, M.-C. Deboin, S. Dupré, A. Menin, M. Winckler and A. Tchechmedjiev. 'Semantic Indexing of Open Scientific Literature to Help Users Discover and Navigate through Publications Networks'. In: *Biodiversity Information Science and Standards, ISSN 2535-0897*. TDWG 2022 - The Biodiversity Information Standards annual conference. Vol. 6. Sofia, Bulgaria, 23rd Aug. 2022. DOI: 10.3897/biss.6.93640. URL: https://hal.mines-ales.fr/hal-03762964.

[75] B. Molinet, S. Marro, E. Cabrio, S. Villata and T. Mayer. 'ACTA 2.0: A Modular Architecture for Multi-Layer Argumentative Analysis of Clinical Trials'. In: IJCAI 2022 - Thirty-First International Joint Conference on Artificial Intelligence. Vienna, Austria, 23rd July 2022. URL: https://hal.inria.fr/hal-03811323.

[76] A. Ollagnier, E. Cabrio, S. Villata and C. Blaya. 'CyberAgressionAdo-v1: a Dataset of Annotated Online Aggressions in French Collected through a Role-playing Game'. In: *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022),* Language Resources and Evaluation Conference. Marseille, France, 20th June 2022. URL: https://hal.science/hal-03765860.

[77]   S. Ren, S. Letz, Y. Orlarey, D. Fober, R. Michon, M. Buffa and L. Pottier. 'Modernized Toolchains to
       Create JSPatcher Objects and WebAudioModules from Faust Code'. In: WAC 2022 - 7th Interna-
       tional Web Audio Conference. Cannes, France, 6th July 2022. DOI: 10.5281/zenodo.6767596.
       URL: https://hal.inria.fr/hal-03812938.

[78]   F. A. S. Robert, M. Winckler, L. Sassatelli and H.-Y. Wu. 'Research Proposal: Analyzing and Under-
       standing Embodied Interactions in Virtual Reality Systems'. In: ACM MMSys 2022 - 13th ACM
       Multimedia Systems Conference. Athlone, Ireland, 14th June 2022. DOI: 10.1145/3524273.3533
       928. URL: https://hal.inria.fr/hal-03698833.

[79]   M. Tikat, A. Menin, M. Buffa and M. Winckler. 'Engineering Annotations to Support Analytical
       Provenance in Visual Exploration Processes'. In: *Lecture Notes in Computer Science*. ICWE 2022 -
       22nd International Conference of Web Engineering. Vol. LNCS - 13362. Web Engineering. Bari,
       Italy, 5th July 2022, pp. 1–16. DOI: 10.1007/978-3-031-09917-5_14. URL: https://hal.scie
       nce/hal-03779349.

[80]   A. Toulet, F. Michel, A. Bobasheva, A. Menin, S. Dupré, M.-C. Deboin, M. Winckler and A. Tchechmed-
       jiev. 'ISSA: Generic Pipeline, Knowledge Model and Visualization tools to Help Scientists Search
       and Make Sense of a Scientific Archive'. In: ISWC 2022 - 21st International Semantic Web Confer-
       ence. Hangzhou, China, 2022. DOI: 10.1007/978-3-031-19433-7_38. URL: https://hal.sci
       ence/hal-03807744.

[81]   V. Vorakitphan, E. Cabrio and S. Villata. 'PROTECT: A Pipeline for Propaganda Detection and
       Classification'. In: CLiC-it 2021- Italian Conference on Computational Linguistics. Milan, Italy,
       26th Jan. 2022. URL: https://hal.science/hal-03417019.

[82]   N. Yacoubi, D. Graux and C. Faron. 'Multi-Level Visual Tours of Weather Linked Data'. In: Pro-
       ceedings of the Seventh International Workshop on the Visualization and Interaction for Ontolo-
       gies and Linked Data (VOILA) co-located with the 21st International Semantic Web Conference.
       Hangzhou, China, 2022. URL: https://hal.inria.fr/hal-03890035.

[83]   N. Yacoubi Ayadi, C. Faron, F. Michel, F. Gandon and O. Corby. 'A Model for Meteorological Knowl-
       edge Graphs: Application to Météo-France Data'. In: ICWE 2022- 22nd International Conference
       on Web Engineering. 22nd International Conference on Web Engineering, ICWE 2022. Bari, Italy,
       5th July 2022. URL: https://hal.inria.fr/hal-03619869.

National peer-reviewed Conferences

[84]   F. Gandon, N. Halliwell and F. Lecue. 'Evaluation d'explications pour la prédiction de liens dans
       les graphes de connaissances par des réseaux convolutifs'. In: Ingénierie des Connaissances 2022
       - IC 2022- PFIA 2022. Saint - Etienne, France, 29th June 2022. URL: https://hal.inria.fr/hal-
       03927901.

[85]   M. Tounsi Dhouib, C. Faron and O. Rodríguez Rocha. 'Comparison of the use of a standard
       repository and an internal repository for the recommendation of learning objects'. In: PFIA
       2022 - Environnements Informatiques pour l'Apprentissage Humain et Intelligence Artificielle.
       Saint-Étienne, France, 2022. URL: https://hal.science/hal-03886251.

[86]   M. Tounsi Dhouib, C. Faron and O. Rodríguez Rocha. 'Learning object recommendation based on
       learning objectives using sentence embeddings models'. In: PFIA 2022 - Conférence Nationale sur
       les Applications Pratiques de l'Intelligence Artificielle 2022. Saint-Étienne, France, 30th June 2022.
       URL: https://hal.science/hal-03662801.

[87]   N. Yacoubi Ayadi, C. Faron, F. Michel, R. Bossy and A. Barbe. 'Construction d'un graphe de
       connaissances à partir des annotations d'articles scientifiques et de leur contenu en sciences de la
       vie'. In: IC'2022 - PFIA 2022 Journées francophones d'Ingénierie des Connaissances. Saint-Etienne,
       France, 29th June 2022. URL: https://hal.inria.fr/hal-03889968.

Scientific book chapters

[88] P.-A. Mattei and S. Villata. 'Introduction à l'intelligence artificielle et aux modèles génératifs'. In: *Informatique Mathématique: Une photographie en 2022*. CNRS Editions, 2022. URL: https://hal.science/hal-03849387.

[89] A. Menin, P. Maillot, C. Faron, O. Corby, C. M. Dal Sasso Freitas, F. Gandon and M. Winckler. 'LDViz: a tool to assist the multidimensional exploration of SPARQL endpoints'. In: *WEBIST 2020/2021, LNBIP 469*. 2023. URL: https://hal.science/hal-03929913.

Doctoral dissertations and habilitation theses

[90] A. E. A. Djebri. 'Uncertainty Management for Linked Data Reliability on the Semantic Web'. Université Côte D'Azur, 24th Feb. 2022. URL: https://hal.science/tel-03679118.

[91] 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.

Reports & preprints

[92] M. Toure, F. Gandon, K. Guidedi, M. Lo and C. Guéret. *Survey of Models and Architectures to Ensure Linked Data Access*. Inria & Université Nice Sophia Antipolis, CNRS, I3S, Sophia Antipolis, France, 23rd Jan. 2023, p. 23. URL: https://hal.inria.fr/hal-03956456.

Other scientific publications

[93] A. Ettorre, F. Michel and C. Faron. 'Prediction of Students' performance in E-learning Environments based on Link Prediction in a Knowledge Graph'. In: AIED 2022 - The 23rd International Conference on Artificial Intelligence in Education. Vol. 13355. Durham, United Kingdom: Springer, 2022. DOI: 10.1007/978-3-031-11647-6_86. URL: https://hal.science/hal-03688838.

[94] N. Halliwell, F. Gandon and F. Lecue. 'A Simplified Benchmark for Ambiguous Explanations of Knowledge Graph Link Prediction using Relational Graph Convolutional Networks'. In: AAAI 2022 - 36th AAAI Conference on Artificial Intelligence. Vancouver, Canada, 22nd Feb. 2022. URL: https://hal.science/hal-03434544.

[95] P. Maillot, O. Corby, C. Faron, F. Gandon and F. Michel. 'KartoGraphI: Drawing a Map of Linked Data'. In: ESWC 2022 - 19th European Semantic Web Conferences. Hersonissos, Greece: Springer, 29th May 2022. URL: https://hal.science/hal-03652865.

[96] N. Yacoubi Ayadi, C. Faron, F. Michel, F. Gandon and O. Corby. 'WeKG-MF: a Knowledge Graph of Observational Weather Data'. In: ESWC 2022 - 19th Extended Semantic Web Conference. Hersonissos, Greece, 29th May 2022. URL: https://hal.inria.fr/hal-03657694.

11.3   Other

Scientific popularization

[97] M. Buffa. 'Audio on the Web (a brief history)'. In: SMC 2022 - Sound and Music Computing. Saint-Etienne, France, 5th June 2022. URL: https://hal.inria.fr/hal-03871527.

[98] F. Gandon. 'Cours toujours'. In: *1024 : Bulletin de la Société Informatique de France* (Nov. 2022), pp. 129–138. DOI: 10.48556/sif.1024.20.129. URL: https://hal.inria.fr/hal-03921374.

[99] A. Toulet, F. Michel and A. Tchechmedjiev. 'Le projet ISSA : l'intelligence artificielle au service de la recherche bibliographique'. In: *Arabesques* 107 (2nd Sept. 2022), pp. 6–7. DOI: 10.35562/arabesques.3046. URL: https://hal.science/hal-03938382.

11.4   Cited publications

[100]   A. Ciortea, S. Mayer, S. Bienz, F. Gandon and O. Corby. 'Autonomous search in a social and ubiquitous Web'. In: *Personal and Ubiquitous Computing* (June 2020). DOI: 10.1007/s00779-02 0-01415-1. URL: https://hal.inria.fr/hal-03135088.

[101]   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. URL: https://hal.i nria.fr/hal-01223236.

[102]   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.