

RESEARCH CENTRE

Sophia Antipolis - Méditerranée

IN PARTNERSHIP WITH:

CNRS, Université Côte d'Azur

2020

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**

# Contents

<b>Project-Team WIMMICS</b>	<b>1</b>
<b>1 Team members, visitors, external collaborators</b>	<b>4</b>
<b>2 Overall objectives</b>	<b>6</b>
2.1 Context and Objectives . . . . .	6
2.2 Research Topics . . . . .	6
<b>3 Research program</b>	<b>7</b>
3.1 Users Modeling and Designing Interaction on the Web and with AI systems . . . . .	7
3.2 Communities and Social Media Interactions and Content Analysis on the Web and Linked Data . . . . .	8
3.3 Vocabularies, Semantic Web and Linked Data Based Knowledge Representation and Extraction of Knowledge Graphs on the Web . . . . .	8
3.4 Artificial Intelligence Processing: Learning, Analyzing and Reasoning on Heterogeneous Knowledge Graphs . . . . .	8
<b>4 Application domains</b>	<b>9</b>
4.1 Social Semantic Web . . . . .	9
4.2 Linked Data on the Web and on Intranets . . . . .	9
4.3 Assisting Web-based Epistemic Communities . . . . .	9
4.4 Linked Data for a Web of Diversity . . . . .	10
4.5 Artificial Web Intelligence . . . . .	10
4.6 Human-Data Interaction (HDI) on the Web . . . . .	10
4.7 Analysis of scientific co-publication . . . . .	11
<b>5 Highlights of the year</b>	<b>11</b>
5.1 Awards . . . . .	12
<b>6 New software and platforms</b>	<b>12</b>
6.1 New software . . . . .	12
6.1.1 CORESE . . . . .	12
6.1.2 DBpedia . . . . .	12
6.1.3 Fuzzy labelling argumentation module . . . . .	13
6.1.4 Qakis . . . . .	13
6.1.5 Corese Server . . . . .	13
6.1.6 CREEP semantic technology . . . . .	13
6.1.7 Licentia . . . . .	14
6.1.8 lod2quiz . . . . .	15
6.1.9 SPARQL micro-services . . . . .	15
6.1.10 ACTA . . . . .	15
6.1.11 WebAudio tube guitar amp sims CLEAN, DISTO and METAL MACHINES . . . . .	16
6.1.12 Morph-xR2RML . . . . .	16
6.1.13 ARViz . . . . .	17
6.1.14 MGExplorer . . . . .	17
<b>7 New results</b>	<b>17</b>
7.1 Users Modeling and Designing Interaction . . . . .	17
7.1.1 LinkedDataViz and MGExplorer . . . . .	17
7.1.2 Visualization of geospatial Linked Data . . . . .	18
7.2 Communities and Social Interactions Analysis . . . . .	18
7.2.1 Autonomous agents in a social and ubiquitous Web . . . . .	18
7.2.2 Multilingual Hate Speech Detection . . . . .	18
7.2.3 Supporting Fake News Identification through Stance Detection . . . . .	19
7.2.4 Aspect-based Sentiment Analysis in Polarized Contexts . . . . .	19

7.2.5	Fuzzy Polarity Propagation for Multi-Domain Sentiment Analysis	20
7.2.6	Linking interactive WebAudio applications to the WASABI knowledge base	20
7.2.7	Using Agent-Based Modeling to explore the role of socio-environmental interactions on Ancient Settlement Dynamics	20
7.3	Vocabularies, Semantic Web and Linked Data Based Knowledge Representation and Artificial Intelligence Formalisms on the Web	20
7.3.1	Publication of the Covid-on-the-Web dataset	20
7.3.2	Mining the Covid-on-the-Web Data	21
7.3.3	Publication of the WASABI dataset	21
7.3.4	Semantic Web for Biodiversity	21
7.3.5	Enriching the WASABI Song Corpus with Lyrics Annotations.	22
7.3.6	Ontology alignment in the sourcing domain	22
7.3.7	A feature-based comparative analysis of legal ontologies	22
7.4	Analyzing and Reasoning on Heterogeneous Semantic Graphs	23
7.4.1	Uncertainty Evaluation for Linked Data	23
7.4.2	Leveraging Data with Uncertain Labels for Machine Learning	23
7.4.3	SPARQL Function: LDScript	23
7.4.4	Linked Data Access and Event Driven programming	23
7.4.5	Linked Data Crawling	24
7.4.6	Semantic Overlay Network for Linked Data Access	24
7.4.7	SHACL Extension	24
7.4.8	Injection of Knowledge in a Sourcing Recommender System	25
7.4.9	Identifying argumentative structures in clinical trials	25
7.4.10	Relation Prediction in Argument Mining	25
7.4.11	Injection of Automatically Selected DBpedia Subjects in Electronic Medical Records to boost Hospitalization Prediction	26
7.4.12	A Knowledge Graph Enhanced Learner Model to Predict Outcomes to Questions	26
7.4.13	Machine Learning for Operations Research	26
7.4.14	RDF Mining	27
<b>8</b>	<b>Bilateral contracts and grants with industry</b>	<b>27</b>
8.1	Bilateral contracts with industry	27
8.2	Bilateral grants with industry	29
<b>9</b>	<b>Partnerships and cooperations</b>	<b>29</b>
9.1	International initiatives	29
9.1.1	Inria associate team not involved in an IIL	29
9.1.2	Participation in other international programs	30
9.2	International research visitors	30
9.2.1	Visits of international scientists	30
9.3	European initiatives	30
9.3.1	FP7 & H2020 Projects	30
9.3.2	Collaborations in European programs, except FP7 and H2020	33
9.4	National initiatives	33
9.5	Regional initiatives	38
<b>10</b>	<b>Dissemination</b>	<b>38</b>
10.1	Promoting scientific activities	38
10.1.1	Scientific events: organisation	38
10.1.2	Scientific events: selection	39
10.1.3	Journal	40
10.1.4	Invited talks	40
10.1.5	Leadership within the scientific community	41
10.1.6	Scientific expertise	41
10.1.7	Research administration	41

10.2 Teaching Supervision Juries . . . . .	41
10.2.1 Teaching . . . . .	41
10.2.2 Supervision . . . . .	43
10.2.3 Juries . . . . .	44
10.2.4 Teaching Administration . . . . .	46
10.3 Popularization . . . . .	46
10.3.1 Articles and contents . . . . .	46
<b>11 Scientific production</b> . . . . .	<b>46</b>
11.1 Major publications . . . . .	46
11.2 Publications of the year . . . . .	47
11.3 Other . . . . .	52
11.4 Cited publications . . . . .	52

## Project-Team WIMMICS

*Creation of the Team: 2012 January 01, updated into 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
- A3.2. – 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
- A4. – Security and privacy
- A4.7. – Access control

- A5.1. – Human-Computer Interaction
  - A5.1.1. – Engineering of interactive systems
  - A5.1.2. – Evaluation of interactive systems
- 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. – Artificial intelligence
  - A9.1. – Knowledge
  - A9.2. – Machine learning
  - A9.4. – Natural language processing
  - A9.5. – Robotics
  - 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
- B5.6. – 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
  - B9.3. – 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]
- Franck Michel [CNRS, Researcher]
- Serena Villata Milanesio [CNRS, Researcher, HDR]

## Faculty Members

- Michel Buffa [Univ Côte d'Azur, Associate Professor, HDR]
- Elena Cabrio [Univ Côte d'Azur, Associate Professor, HDR]
- Catherine Faron [Univ Côte d'Azur, Associate Professor, HDR]
- Clement Jonquet [Univ Montpellier II (sciences et techniques du Languedoc), Associate Professor, until Aug 2020, HDR]
- Nhan Le Thanh [Univ Côte d'Azur, Professor]
- Peter Sander [Univ Côte d'Azur, Professor]
- Andrea Tettamanzi [Univ Côte d'Azur, Professor, HDR]
- Marco Winckler [Univ Côte d'Azur, Professor, from Feb 2020, HDR]

## Post-Doctoral Fellows

- Jerome Delobelle [Inria, until Aug 2020]
- Raphaël Gazzotti [Inria, from Nov 2020]
- Aline Menin [Inria, from Dec 2020]
- Iliana Petrova [Inria, from May 2020]
- Stefan Sarkadi [Inria, from Nov 2020]

## PhD Students

- Ali Ballout [Univ Côte d'Azur, from Oct 2020]
- Lucie Cadorel [Inria, from May 2020]
- Dupuy Rony Charles [KINAXIA Company, CIFRE, from Sep 2020]
- Molka Dhouib [SILEX Company, CIFRE]
- Ahmed Elamine Djebri [Algerian Government]
- Antonia Ettore [Univ Côte d'Azur]
- Michael Fell [CNRS, until May 2020]
- Nicholas Halliwell [Inria]
- Mina Ayse Ilhan [Univ Côte d'Azur]
- Adnane Mansour [Ecole Nationale Supérieure des Mines de Saint Etienne, from Dec 2020]



- Santiago Marro [Univ Côte d'Azur, from Apr 2020]
- Tobias Mayer [Univ Côte d'Azur]
- Thu Huong Nguyen [Ministry of Education of Vietnam]
- Shihong Ren [St Etienne University, from Dec 2020]
- Maroua Tikat [Univ Côte d'Azur, from Oct 2020]
- Mahamadou Toure [UGB Sénégal]
- Vorakit Vorakitphan [Inria]

### **Technical Staff**

- Anna Bobasheva [Inria, Engineer]
- Erwan Demairy [Inria, Engineer]
- Raphaël Gazzotti [Inria, Engineer, from May 2020 until Oct 2020]
- Hai Huang [Inria, Engineer]
- Aline Menin [Inria, Engineer, from Jun 2020 until Nov 2020]

### **Interns and Apprentices**

- Valentin Ah-Kane [Univ Côte d'Azur, from Apr 2020 until Sep 2020]
- Valeria Bellusci [University of InsubriaItaly, from Mar 2020 until Aug 2020]
- Dorian Chapoulie [CNRS, from May 2020 until Aug 2020]
- Jean Marie Dormoy [Inria, Intern, from Jun 2020 until Aug 2020]
- Mathis Le Quiniou [Inria, from Sep 2020]
- Abdelhadi Lebbar [Inria, from Mar 2020 until Aug 2020]
- Benjamin Molinet [Inria, Apprentice, from Oct 2020]
- Mohamed Amine Romdhane [Inria, from Jun 2020 until Aug 2020]
- Yuting Sun [Inria, from Mar 2020 until Aug 2020]
- Maroua Tikat [Univ Côte d'Azur, from Mar 2020 until Aug 2020]

### **Administrative Assistant**

- Christine Foggia [Inria]

### **Visiting Scientists**

- Yimin Hu [Chinese Academy of Science]
- Dario Malchiodi [University of Milan Italy , from Oct 2020]
- Yuting Sun [Univ Côte d'Azur , until Jan 2020]

## External Collaborators

- Andrei Ciortea [Univ St Gallen Switzerland]
- Claude Frasson [Montreal University Canada, until Mar 2020, HDR]
- Raphaël Gazzotti [Synchronext Company , until Apr 2020]
- Alain Giboin [Self-employed]
- Freddy Lecue [Thalès]
- Oscar Rodríguez Rocha [TeachOnMars Company]

## 2 Overall objectives

### 2.1 Context and Objectives

The Web became a virtual place where persons and software interact in mixed communities. The Web has the potential of becoming the collaborative space for natural and artificial intelligence, raising the problem of supporting these worldwide interactions. These large scale mixed interactions create many problems that must be addressed with multidisciplinary approaches [67].

One particular problem is to reconcile formal semantics of computer science (e.g. logics, ontologies, typing systems, protocols, etc.) on which the Web architecture is built, with soft semantics of people (e.g. posts, tags, status, relationships, etc.) on which the Web content is built.

*Wimmics proposes models and methods to bridge formal semantics and social semantics on the Web [68] in order to address some of the challenges in building a Web as a universal space linking many different kinds 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. And 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 the means to combine them and to perform multi-criteria analysis on their combination. Wimmics contributes to this understanding 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 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 meaningful visual representations?" "How to guide the analysis of graphs that might contain data with diverse levels of accuracy, precision and interestingness to the users?"

**Topic 2 - communities and social interactions and content analysis on the Web:** 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 them in order for software to exploit these knowledge graphs in their processing (in the next objective).

**Topic 4 - artificial intelligence processing: learning, analyzing and reasoning on heterogeneous 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: SMILK Joint Laboratory, Corese, DBpedia.fr.

### 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: EduMICS, OCKTOPUS, Vigiglobe, Educlever, Gayatech.

#### **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: Zoomathia, Seempad, SMILK,

#### **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: MoreWAIS, Corese, Vigiglobe collaboration.

#### **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: QAKIS, Sychonext collaboration, ALOOF, DiscoveryHub, WASABI, MoreWAIS.

**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.

#### 4.7 Analysis of scientific co-publication

Over the last decades, scientific research has matured and diversified. In all areas of knowledge, we observe an increasing number of scientific publication, a rapid development of more and more specialized conferences and journals, and the creation of dynamic collaborative networks that cross borders and evolve over time. In this context, the analysis of scientific publications becomes a major issue for the sustainability of scientific research. To illustrate this, let's 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 dataset 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 relevant information.

By reasoning on the Linked Data and Web semantic schemas, we investigate methods and tools enabling users to find relevant publications. Hereafter we present some example of typical problems for the analysis of co-publications 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 full 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 of argument graphs that describe association and development of ideas in scientific papers.
- How to understand the collaboration of authors in the development of scientific knowledge? For that, we have used 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 data, and the Covid-on-the-Web datasets.

## 5 Highlights of the year

As soon as the Covid crisis put France in lockdown in March 2020, the team started the project Covi-dOnTheWeb to allow biomedical researchers to access, query and make sense of COVID-19 scholarly

literature [48, 21].

Damien Graux was recruited as a new tenured junior researcher for the team <https://dgraux.github.io/>.

HDR Defense of Elena Cabrio [57]

Publication of the third edition of the textbook “Semantic Web for the Working Ontologist” [51] with Fabien Gandon as new co-author.

## 5.1 Awards

Elena Cabrio, Serena Villata, Michel Buffa and Fabien Gandon received Université Côte d’Azur medals for their work in 2020.

# 6 New software and platforms

## 6.1 New software

### 6.1.1 CORESE

**Name:** COnceptual REsource Search Engine

**Keywords:** Semantic Web, Search Engine, RDE, SPARQL

**Functional Description:** Corese is a Semantic Web Factory, it implements W3C RDE, 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

### 6.1.2 DBpedia

**Name:** DBpedia

**Keywords:** RDE, SPARQL

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

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

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

**Contact:** Fabien Gandon

**Participants:** Fabien Gandon, Elmahdi Korfed



### 6.1.3 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

### 6.1.4 Qakis

**Name:** Question-Answering wiki framework based system

**Keyword:** Natural language

**Functional Description:** The QAKiS system implements question answering over DBpedia. QAKiS allows end users to submit a query to an RDF triple store in English and to obtain the answer in the same language, hiding the complexity of the non-intuitive formal query languages involved in the resolution process. At the same time, the expressiveness of these standards is exploited to scale to the huge amounts of available semantic data. Its major novelty is to implement a relation-based match for question interpretation, to convert the user question into a query language (e.g. SPARQL). English, French and German DBpedia chapters are the RDF data sets to be queried using a natural language interface.

**URL:** <http://www.qakis.org/>

**Contact:** Elena Cabrio

**Participants:** Alessio Palmero Aprosio, Amine Hallili, Elena Cabrio, Fabien Gandon, Julien Cojan, Serena Villata

### 6.1.5 Corese Server

**Name:** Corese Server

**Keywords:** Semantic Web, RDE, SPARQL

**Scientific Description:** A Web server to interact with Corese via HTTP SPARQL endpoint, STTL display engine

**Contact:** Olivier Corby

**Participants:** Alban Gaignard, Fuqi Song, Olivier Corby

**Partner:** I3S

### 6.1.6 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 network 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

### 6.1.7 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

### 6.1.8 lod2quiz

**Keywords:** Semantic Web, Artificial intelligence, Web Application, E-learning

**Functional Description:** Automatic quiz generator

**Release Contributions:** Contains the core engine of lod2quiz, deployed as a web application that exposes a REST API with methods for the quiz generation.

**Publications:** [hal-01688798](#), [hal-01811490](#), [hal-01758737](#)

**Contact:** Oscar Rodriguez Rocha

**Participants:** Oscar Rodriguez Rocha, Catherine Faron

### 6.1.9 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

### 6.1.10 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

### 6.1.11 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

### 6.1.12 Morph-xR2RML

**Name:** Morph-xR2RML

**Keywords:** RDE, 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 RDE. 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

### 6.1.13 ARViz

**Name:** Association Rules Visualization

**Keyword:** Information visualization

**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/>

**Contact:** Marco Antonio Alba Winckler

### 6.1.14 MGExplorer

**Name:** Multivariate Graph Explorer

**Keyword:** Information visualization

**Functional Description:** MGExplorer is an information visualization toolsuit 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.

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

**URL:** <http://covid19.i3s.unice.fr:8080/>

**Contact:** Marco Antonio Alba Winckler

**Partner:** Universidade Federal do Rio Grande do Sul

## 7 New results

### 7.1 Users Modeling and Designing Interaction

#### 7.1.1 LinkedDataViz and MGExplorer

<b>Participants</b>	Marco Winckler, Aline Menin, Olivier Corby, Alain Giboin, Fabien Gandon.
---------------------	--

Visualization techniques are useful tools to explore datasets, enabling the discovery of meaningful patterns and causal relationships. Nonetheless, the discovery process is often exploratory and requires multiple views to support analyzing different or complementary perspectives to the data. The analytical reasoning that guides the exploration processes based on multiple views can be represented by provenance between views. In this paper [71], we introduce the term ancillary search tasks to characterize multiple complementary search tasks (possibly run in parallel) that help users to achieve a complex search task. This concept has been extended to support chained views to describe the incremental exploration of large, multidimensional datasets through the combination of multiple chained visualization techniques and visual querying, and the representation of analytical provenance through a visual representation of the dependencies between views. As a proof-of-concept, we developed a visualization tool MGExplorer, which encompasses a sample of five visualization techniques (Node-Edge Diagram, ClusterVis, GlyphMatrix, Histogram, and IRIS) that are used to explore multivariate graphs datasets. Each view in MGExplorer

supports visual querying techniques that enable the definition of subsets of the current dataset to be explored in another, chained view.

Linked Data Viz is a platform to provide graphic views of Linked Data. The platform is now generic in the sense that it can query any SPARQL endpoint. A specific service has been designed in order to submit a SPARQL query and the URL of a SPARQL endpoint. The outcomes of Linked Data Viz are used as entry point for visualizations created by the tool MGExplorer.

LinkedDataViz web site: <http://covid19.i3s.unice.fr:8080/ldviz>.

### 7.1.2 Visualization of geospatial Linked Data

**Participants** Franck Michel, Marco Winckler, Olivier Corby.

By means of an Ubinet Master internship, we initiated a work meant to explore the cross-fertilization of geospatial data visualization and reasoning on linked data. This nascent work sheds light on interesting leads that we intend to push further.

## 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 [9].

### 7.2.2 Multilingual Hate Speech Detection

**Participants** Elena Cabrio, Serena Villata, Michele Corazza.

The increasing popularity of social media platforms like Twitter and Facebook has led to a rise in the presence of hate and aggressive speech on these platforms. Despite the number of approaches recently proposed in the Natural Language Processing research area for detecting these forms of abusive language, the issue of identifying hate speech at scale is still an unsolved problem. In this research activity, together with Sara Tonelli (FBK Trento) and Stefano Menini (FBK Trento), we have proposed a robust recurrent neural architecture which is shown to perform in a satisfactory way across different languages, namely English, Italian and German. We address an extensive analysis of the obtained experimental results over the three languages to gain a better understanding of the contribution of the different components

employed in the system, both from the architecture point of view (i.e., Long Short Term Memory, Gated Recurrent Unit, and bidirectional Long Short Term Memory) and from the feature selection point of view (i.e., social network specific features, emotion lexica, emojis, embeddings). To address such in-depth analysis, we use three freely available datasets for hate speech detection on social media on English, Italian and German [10].

### 7.2.3 Supporting Fake News Identification through Stance Detection

**Participants** Elena Cabrio, Serena Villata, Jérôme Delobelle.

This work is part of the DGA project RAPID CONFIRMA (COntre argumentation contre les Fausses InfoRMAtion) aiming to automatically detect fake news and limit their diffusion. In our work, we present a concrete application scenario where a fake news detection system is empowered with an argument mining model, to highlight and aid the analysis of the arguments put forward to support or oppose a given target topic in articles containing fake information [41]. More precisely, we propose to extend a disinformation analysis tool with a stance detection module for arguments relying on pretrained language models, i.e., BERT, with the aim of obtaining a more effective analysis tool both for users and analysts. To evaluate the argument stance detection module in the disinformation context, we propose to annotate a new resource of fake news articles, where arguments are classified as being *In Favor* or *Against* towards a target topic. Our new annotated data set contains sentences about three topics currently attracting a lot of fake news around them, i.e., public health demands vaccination, white helmets provide essential services, and the risible impact of Covid-19. This data set collects 86 articles containing nearly 3000 sentences.

### 7.2.4 Aspect-based Sentiment Analysis in Polarized Contexts

**Participants** Vorakit Vorakitphan, Elena Cabrio, Serena Villata.

Aspect-based Sentiment Analysis (ABSA) aims at capturing sentiment (i.e., positive, negative or neutral) expressed toward each aspect (i.e., attribute) of a target entity. The main interest is to capture sentiment nuances about different entities. However, in a context of opinion polarization, different groups of people can form strong convictions of competing opinions on such target entities, resulting in different (often opposite) evaluations of the same aspect. Compare, for example, the differences in the pro- and anti-Brexit discourses concerning the withdrawal of the United Kingdom from the European Union, aligning with contrasting attitudes toward the EU, the immigration and the country's culture. Whilst in standard scenarios of sentiment analysis about specific entities and their aspects it is assumed that sentiment is consistent (e.g., a big screen is a desirable characteristics for a TV), this is not the case for polarized contexts. Hence, for example, a "clean Brexit" might be desirable to some, but not to others. Together with Marco Guerini (FBK Trento), we proposed a comprehensive framework for studying the interaction of ABSA with opinion polarization in newspapers and social media [38]. We first trained a machine learning algorithm that detects the emotions and their intensities at sentence-level, and then we mapped emotion intensities into the Valence, Arousal, and Dominance (VAD) model. Later, we built a framework to assess whether and how VAD are connected to polarized contexts, by computing the VAD scores of a set of key-concepts that can be found on newspapers with opposite views. These key-concepts (e.g., "stop immigration") are built from a set of aspects ("immigration" in our example) combined with relevant verbs or adjectives that represent a clear polarized opinion toward the aspect (e.g., "stop"). To experiment with the proposed approach, we focussed on the Brexit scenario as it provided us with the required elements to carry out our study, because of the opinion divisions formed around one or more political positions or issues. In our experimental setting, we selected two British newspapers known to be polarized, i.e., either for or against Brexit. Results show that VAD are not absolute, but relative to the newspaper's viewpoint on the key-concept. Our approach highlights that using the proposed

key-concepts gives us fine-grained details about VAD elements that strongly interact with the polarized context. We showed that standard SA approaches can be deceptive in such polarized setting (considering only the word "Brexit" on both newspapers, the valence is almost identical), while our ABSA approach showed a clear-cut polarization.

### 7.2.5 Fuzzy Polarity Propagation for Multi-Domain Sentiment Analysis

**Participants** Andrea Tettamanzi.

Together with Claude Pasquier and Célia da Costa Pereira of the I3S Laboratory, we studied how different domain-dependent polarities can be learned for the same concepts, in the context of multi-domain sentiment analysis. To this aim, we extend an existing approach based on the propagation of fuzzy polarities over a semantic graph capturing background linguistic knowledge to learn concept polarities with respect to various domains and their uncertainty from labeled datasets. In particular, we use POS tagging to refine the association between terms and concepts and word embedding to enhance the construction of the semantic graph. The proposed approach [34] was then evaluated on a standard benchmark, showing that the combined use of POS tagging and word embedding improves its performance. One particularly strong point of the proposed approach is its recall, which is always very close to 100%. In addition, it exhibits good cross-domain generalization capabilities.

### 7.2.6 Linking interactive WebAudio applications to the WASABI knowledge base

**Participants** Michel Buffa.

In the context of the WASABI research project, we built a 2M song database made of metadata collected from the Web of Data and from the analysis of song lyrics [44] of the audio files provided by Deezer (and sometimes from other sources such as YouTube [54]). We designed a WebAudio plugin standard, new tools for developing high performances plugins in the browser [14], and new methods for real-time tube guitar amplifier simulations that run in the browser [19]. Some of these results are unique in the world as in 2020, and have been acclaimed by two 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.com](http://ampedstudio.com) [20]. 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 [35].

### 7.2.7 Using Agent-Based Modeling to explore the role of socio-environmental interactions on Ancient Settlement Dynamics

**Participants** Andrea Tettamanzi.

Within the framework of a multi-disciplinary project involving archaeologists, economists, geographers, and computer scientists, we used Agent-Based Modelling to explore the respective impacts of environmental and social factors on the settlement pattern and dynamics during the Roman period in South-Eastern France [52].

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

### 7.3.1 Publication of the Covid-on-the-Web dataset



**Participants** Franck Michel, Fabien Gandon, Valentin Ah-Kane, Anna Bobasheva, Elena Cabrio, Olivier Corby, Raphaël Gazzotti, Alain Giboin, Santiago Marro, Tobias Mayer, Serena Villata, Marco Winckler.

The Covid-on-the-Web project aims to allow biomedical researchers to access, query and make sense of COVID-19 related literature. Launched in Mars 2020, it involved multiple skills of the team in knowledge representation, text, data and argument mining, as well as data visualization and exploration. Among the achievements of the projetct is the Covid-on-the-Web RDF dataset [48] that we genetared and published by processing, analyzing and enriching the “COVID-19 Open Research Dataset” (CORD-19) that gathers 100K+ full-text scientific articles related to the coronaviruses. The dataset produced comprises two main knowledge graphs: (1) named entities 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.

Web site <https://github.com/Wimmics/CovidOnTheWeb>.

### 7.3.2 Mining the Covid-on-the-Web Data

**Participants** Lucie Cadorel, Andrea Tettamanzi.

As soon as the Covid-on-theWeb RDF dataset was published, we set out to exploit it to mine interesting associations from it. We thus proposed a method to discover interesting association rules from an RDF knowledge graph, by combining clustering, community detection, and dimensionality reduction, as well as criteria for filtering the discovered association rules in order to keep only the most interesting rules [21]. Our results demonstrate the effectiveness and scalability of the proposed method and suggest several possible uses of the discovered rules, including (i) curating the knowledge graph by detecting errors, (ii) finding relevant and coherent collections of scientific articles, and (iii) suggesting novel hypotheses to biomedical researchers for further investigation.

### 7.3.3 Publication of the WASABI dataset

**Participants** Franck Michel, Fabien Gandon, Elena Cabrio, Alain Giboin, Marco Winckler, Maroua Tikat, Michael Fell.

Since 2017, a two-million song database consisting of metadata collected from multiple open data sources and automatically extracted information has been constructed in the context of the WASABI project. The goal is to build a knowledge graph linking collected metadata (artists, discography, producers, dates, etc.) with metadata generated by the analysis of both the songs’ lyrics (topics, places, emotions, structure, etc.) and audio signal (chords, sound, etc.). It relies on natural language processing and machine learning methods for extraction, and semantic Web frameworks for integration. The dataset describes more than 2 millions commercial songs, 200K albums and 77K artists. It can be exploited by music search engines, music professionals or scientists willing to analyze popular music published since 1950. It is available under an open license in multiple formats and is accompanied by online applications and open source software including an interactive navigator, a REST API and a SPARQL endpoint.

Web site <https://github.com/micbuffa/WasabiDataset>.

### 7.3.4 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 two main axes.

First, in 2019 the MNHN started using our SPARQL Micro-Services architecture and framework to help biologists in editing taxonomic information by confronting multiple, heterogeneous data sources [70]. In 2020 this collaboration has been strengthened and the MNHN now heavily relies on those services for daily activities.

Second, we have kept on the work initiated within the Bioschemas.org W3C community group that seeks the definition and adoption of common biology-related markup terms. While a new term `TaxonName` was defined and we updated MNHN webpages accordingly, we have undertaken an "evangelization" action to promote this practice in the biodiversity community [30].

### 7.3.5 Enriching 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.

### 7.3.6 Ontology alignment in the sourcing domain

**Participants** Molka Dhoub, Catherine Faron, Andrea Tettamanzi.

In the framework of a collaborative project with Silex France company aiming to propose a decision support to recommend relevant providers for a service request, we developed during the last two years a domain knowledge modeling specific to the sourcing domain with the goal of reasoning on knowledge to improve the providers' recommender. We proposed a new ontology alignment approach based on a set of rules exploiting the embedded space and measuring clusters of labels to discover the relationship between concepts. We evaluated our approach on several open datasets from the Ontology Alignment Evaluation Initiative (OAEI) benchmark and a real-world case study provided by the Silex company. This year, we extended our evaluation on another real-world case study provided by the "Office National d'Information sur les Enseignements et les Professions" (ONISEP).

### 7.3.7 A feature-based comparative analysis of legal ontologies

**Participants** Serena Villata.

Ontologies represent the standard way to model the knowledge about specific domains. This holds also for the legal domain where several ontologies have been put forward to model specific kinds of legal knowledge. Both for standard users and for law scholars, it is often difficult to have an overall view on the existing alternatives, their main features and their interlinking with the other ontologies. To answer this need, in this work, we address an analysis of the state-of-the-art in legal ontologies and we characterise

them along with some distinctive features. This work aims to guide generic users and law experts in selecting the legal ontology that better fits their needs and in understanding its specificity so that proper extensions to the selected model could be investigated [13].

## 7.4 Analyzing and Reasoning on Heterogeneous Semantic Graphs

### 7.4.1 Uncertainty Evaluation for Linked Data

**Participant** Ahmed Elamine Djebri, Fabien Gandon, Andrea Tettamanzi.

For data sources to ensure providing reliable linked data, they need to indicate information about the (un)certainly 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 [43].

### 7.4.2 Leveraging Data with Uncertain Labels for Machine Learning

**Participants** Andrea Tettamanzi.

Prompted by an application in the area of human geography using machine learning to study housing market valuation based on the urban form, we proposed a method based on possibility theory to deal with sparse data, which can be combined with any machine learning method to approach weakly supervised learning problems [54]. More specifically, the solution we propose constructs a possibilistic loss function to account for an uncertain supervisory signal. Although the proposal was motivated by a specific application, its basic principles are general. The proposed method has then been empirically validated on real-world data.

### 7.4.3 SPARQL Function: LDScript

**Participant** Olivier Corby.

We have continued the implementation and validation of LDScript, Linked Data Script, a programming language compatible with SPARQL that enables users to write extension functions that are directly executable in SPARQL queries. LDScript is an extension of SPARQL Filter language with function definition, variable declaration, iteration, second order and anonymous function, pattern matching. It provides users with extension datatypes that enables them to manage Semantic Web objects such as RDF triple and graph as well as SPARQL Query result. In addition, extension datatypes provide implementations for list, hashmap, XML document and JSON object. A SHACL interpreter has been entirely written using LDScript.

Linked Data Script documentation: <https://files.inria.fr/corese/doc/ldscript.html>

### 7.4.4 Linked Data Access and Event Driven programming

**Participant** Olivier Corby.

We started a preliminary work on an access control model for RDF graphs. It is possible to specify access rights at the scale of nodes and predicates URIs or namespaces. Preliminary report: <https://files.inria.fr/corese/doc/access.html>.

We started a preliminary work on a safety model to protect a SPARQL endpoint where functions can be protected (forbidden) e.g. Linked Functions. Preliminary report: <https://files.inria.fr/corese/doc/level.html>

We generalized the Event Driven programming model for the HTTP server (SPARQL endpoint), SPARQL Update, SHACL interpreter, Rule engine and Transformation engine.

Linked Data Event Driven Programming documentation: <https://files.inria.fr/corese/doc/event.html>.

#### 7.4.5 Linked Data Crawling

**Participant** Fabien Gandon, Hai Huang.

A Linked Data crawler performs a selection to focus on collecting linked RDF (including RDFa) data on the Web. From the perspectives of throughput and coverage, given a newly discovered and targeted URI, the key issue of Linked Data crawlers is to decide whether this URI is likely to dereference into an RDF data source and therefore if it is worthy downloading the representation it points to. Current solutions adopt heuristic rules to filter irrelevant URIs. But when the heuristics are too restrictive this hampers the coverage of crawling. We proposed and compared approaches to learn strategies for crawling Linked Data on the Web by predicting whether a newly discovered URI will lead to an RDF data source or not. We detailed the features used in predicting the relevance and the methods we evaluated including a promising adaptation of FTRL-proximal online learning algorithm. We compared several options through extensive experiments including existing crawlers as baseline methods to evaluate their efficiency [26].

#### 7.4.6 Semantic Overlay Network for Linked Data Access

**Participant** Fabien Gandon, Mahamadou Toure.

We proposed and evaluated MoRAI (Mobile Read Access in Intermittent internet connectivity), a distributed peer-to-peer architecture organized in three levels dedicated to RDF data exchanges by mobile contributors. We presented the conceptual and technical aspects of this architecture as well as a theoretical analysis of the different characteristics. We then evaluated it experimentally and results show the relevance of considering geographical positions during data exchanges and of integrating RDF graph replication to ensure data availability in terms of requests completion rate and resistance to crash scenarios [37].

#### 7.4.7 SHACL Extension

**Participant** Olivier Corby, Iliana Petrova, Fabien Gandon.

In the context of a collaboration with Stanford University, we have been working on extensions of W3C SHACL Shape Constraint Language<sup>1</sup>.

<sup>1</sup>see <https://www.w3.org/TR/shacl/>

We have proposed extensions of SHACL path language with a `xsh:predicatePath` operator that enables the interpreter to navigate from a node in the RDF graph to the set of predicates the node is subject, object or both. In addition, we propose to extend the path language to navigate from nodes to triples and back with two operators: `xsh:triplePath` and `xsh:nodePath`. The path language is also extended with `xsh:exist` and `xsh:filter` that enable the interpreter to check conditions.

SHACL shape constraints are extended with a `xsh:function` statement that enable the user to specify constraints using LDScript functions. Additional detailed validation results can be obtained for node and boolean constraints.

Linked Data SHACL Extension documentation: <https://files.inria.fr/corese/doc/shacl.html>

#### 7.4.8 Injection of Knowledge in a Sourcing Recommender System

**Participants** Molka Dhouib, Catherine Faron, Andrea Tettamanzi.

In the framework of a collaborative project with Silex France company aiming to propose a decision support to recommend relevant providers for a service request, we proposed a new named entity recognition algorithm combining several types of extracted features describing the textual description of service requests and service providers, such as: (i) semantics, (ii) syntax, (iii) word characters, and (iv) position of words. We use it to construct a vector representation of service requests and service providers. [39] Secondly, we proposed a recommender system approach based on the definition of a similarity measure between the vector representations of service requests and service providers. [36].

#### 7.4.9 Identifying argumentative structures in clinical trials

**Participants** Elena Cabrio, Serena Villata, Tobias Mayer, Santiago Marro.

We annotated first a dataset of 159 abstracts of Randomized Controlled Trials (RCTs) from the MEDLINE database, comprising 4 different diseases (i.e., glaucoma, hypertension, hepatitis b, diabetes), then a larger dataset of 500 abstracts on the neoplasm disease, leading to a dataset of 4113 argument components and 2601 argument relations. We then proposed a complete argument mining pipeline for RCTs, classifying argument components as evidence and claims, and predicting the relation, i.e., attack or support, holding between those argument components [46]. We experimented with deep bidirectional transformers in combination with different neural architectures (i.e., LSTM, GRU and CRF) and outperformed current state-of-the-art end-to-end argument mining systems. In addition, we also included the identification of PICO elements in the abstracts (PICO is a framework to answer health-care related questions in evidence-based practice. Elements comprise patients/population (P), intervention (I), control/comparison (C) and outcome (O) information). We finally investigated the robustness of language models like BERT for the argument classification task [47].

#### 7.4.10 Relation Prediction in Argument Mining

**Participants** Elena Cabrio, Serena Villata.

Argument(ation) Mining (AM) is the research area which aims at extracting argument components and predicting argumentative relations (i.e., support and attack) from text. In particular, numerous approaches have been proposed in the literature to predict the relations holding between arguments and application-specific annotated resources were built for this purpose. Despite the fact that these resources were created to experiment on the same task, the definition of a single relation prediction method to be successfully applied to a significant portion of these datasets is an open research problem

in AM. This means that none of the methods proposed in the literature can be easily ported from one resource to another. Together with Oana Cocarascu and Francesca Toni from Imperial College London (UK), we addressed this problem by proposing a set of dataset independent strong neural baselines which obtain homogeneous results on all the datasets proposed in the literature for the argumentative relation prediction task in AM [22]. Thus, our baselines can be employed by the AM community to compare more effectively how well a method performs on the argumentative relation prediction task.

#### 7.4.11 Injection of Automatically Selected DBpedia Subjects in Electronic Medical Records to boost Hospitalization Prediction

**Participants** Catherine Faron, Fabien Gandon, Raphaël Gazzotti.

Although there are many medical standard vocabularies available, it remains challenging to properly identify domain concepts in electronic medical records. Variations in the annotations of these texts in terms of coverage and abstraction may be due to the chosen annotation methods and the knowledge graphs, and may lead to very different performances in the automated processing of these annotations. We proposed a semi-supervised approach based on DBpedia to extract medical subjects from EMRs and evaluate the impact of augmenting the features used to represent EMRs with these subjects in the task of predicting hospitalization. We compared the impact of subjects selected by experts vs. by machine learning methods through feature selection. Our approach was experimented on data from the database PRIMEGE PACA that contains more than 600,000 consultations carried out by 17 general practitioners (GPs) [25].

#### 7.4.12 A Knowledge Graph Enhanced Learner Model to Predict Outcomes to Questions

**Participants** Antonia Ettore, Catherine Faron, Fabien Gandon, Mathis Le Quiniou, Franck Michel, Oscar Rocha Rodriguez, Yuting Sun.

In order for a learning platform to provide personalized services, the knowledge and skills progressively acquired by students on each subject should be taken into account when choosing the training and evaluation questions to be presented to them, in the form of customized quizzes. To achieve such recommendation, a first step lies in the ability to predict the outcome of students when answering questions (success or failure). We proposed a model of the students' learning able to make such predictions on the SIDES platform for medical students. The model extends a state-of-the-art approach to fit the specificity of medical data, and to take into account additional knowledge extracted from the SIDES knowledge graph in the form of graph embeddings. Through an evaluation based on learning traces for pediatrics and cardiovascular specialties, we showed that considering the vector representations of answers, questions and students nodes substantially improves the prediction results compared to baseline models [24]. In the continuation of this work, we conducted preliminary experiments to test the applicability of our model in other learning environments, namely the TeachOnMars learning platform for in-company training and the Educlever platform for secondary education.

#### 7.4.13 Machine Learning for Operations Research

**Participants** Andrea Tettamanzi.

Together with Alberto Ceselli and Saverio Basso of the University of Milan we used machine learning techniques to understand good decompositions of linear programming problems [6].

#### 7.4.14 RDF Mining

**Participants** Thu Huong Nguyen, Andrea Tettamanzi.

In the framework of Nguyen Thu Huong's thesis, we have continued to explore the use of grammar-based evolutionary method to mine RDF datasets for OWL class disjointness axioms. In particular, we addressed the problem of discovering disjointness axioms involving complex class expressions [32, 33]. As it turns out, this problem involves at least two conflicting criteria an axiom should meet, namely possibility (i.e., truth, acceptability, likelihood), and generality. This prompted us to adapt our evolutionary approach to suit multi-objective optimization [31].

On the other hand, our evolutionary approach critically rely 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. Together with Dario Malchiodi of the University of Milan and Célia da Costa Pereira of the I3S Laboratory, we assess the role of similarity measures and learning methods in classifying candidate axioms for automated schema induction through kernel-based learning algorithms. The evaluation was based on three different similarity measures between axioms and two alternative dimensionality reduction techniques to check the extent to which the considered similarities allow to separate true axioms from false axioms. The result of the dimensionality reduction process is subsequently fed to several learning algorithms, comparing the accuracy of all combinations of similarity, dimensionality reduction technique, and classification method. As a result, it is observed that it is not necessary to use sophisticated semantics-based similarity measures to obtain accurate predictions, and furthermore that classification performance only marginally depends on the choice of the learning method. Our results open the way to implementing efficient surrogate models for axiom scoring to speed up ontology learning and schema induction methods [29].

## 8 Bilateral contracts and grants with industry

### 8.1 Bilateral contracts with industry

#### PREMISSE Collaborative Project

**Participants** Molka Dhouib, Catherine Faron, Andrea Tettamanzi.

Partner: SILEX France.

This collaborative project with the SILEX France company started in March 2017, funded by the ANRT (CIFRE PhD). SILEX France is developing a B2B platform where service providers and consumers upload their service offers or requests in free natural language; the platform is intended to recommend service providers to the applicant, which are likely to fit his/her service request. The aim of this project is to propose a decision support system by exploiting the semantic knowledge that are extracted from the textual descriptions of requests for services and providers, in order to recommend relevant providers for a service request.

#### HealthPredict Collaborative Project

**Participants** Raphaël Gazzotti, Catherine Faron, Fabien Gandon.

Partner: Synchronext.

This collaborative project with the Synchronext company started in April 2017, funded by the ANRT (CIFRE PhD). Synchronext is a startup aiming at developing Semantic Web business solutions. The aim of this project is to design a digital health solution for the early management of patients through consultations with their general practitioner and health care circuit. The goal is to develop a predictive Artificial Intelligence interface that allows to cross the data of symptoms, diagnosis and medical treatments of the population in real time to predict the hospitalization of a patient. We presented at SAC 2020 [25] a semi-supervised approach based on DBpedia to select from electronic medical records subjects designating medical aspects relevant to the prediction of hospitalization.

### Curiosity Collaborative Project

**Participants** Catherine Faron, Oscar Rodríguez Rocha.

Partner: TeachOnMars.

This collaborative project with the TeachOnMars company started in October 2019. TeachOnMars is developing 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., as explained in the section on the context of the project. 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.

## 8.2 Bilateral grants with industry

Accenture gifts (June 2017 - January 2022): Wimmics has received two gifts from Accenture. Together with additional funds from another project these gifts have been used to fund the Engineer position and then the PhD Grant (June 2017 - January 2022) of Nicholas Halliwell on a topic agreed with Accenture: “interpretable and explainable predictions”

# 9 Partnerships and cooperations

## 9.1 International initiatives

### 9.1.1 Inria associate team not involved in an ILL

#### PROTEMICS, SHACL-S and CoP4Pro

**Title:** PROTEMICS

**Duration:** 2020 - 2023

**Coordinator:** Fabien Gandon

**Partners:**

- School of Computing, Stanford (United States)

**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 proposed 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. PROTEMICS is an associate team, SHACL-S is an Exploratory Action (AEx) and CoP4Pro is a Development Action (ADT) and these three complementary projects are addressing the research, collaboration and development aspects of the same topic.

### 9.1.2 Participation in other international programs

#### NOMOS

**Title:** A Model-Based Approach for Designing Territorial User Interfaces

**duration :** 2020-2021

**Coordinator :** Marco Winckler (France) and Jean Vanderdonckt (Belgium)

**partners :** Université Côte d'Azur and Universté catholique de Louvain-la-Neuve

**Contact :** Marco Winckler

**Summary :** NOMOS (the French acronym for Nouvelle Organisation de Modèles Orientés Surfaces pour la conception de systèmes de systèmes interactifs basés sur la territorialité) is an international cooperation project funded by the program Tournesol. The research questions of NOMOS are articulated around the developement of a model-based approach for designing graphical user interfaces that are delineated based on the concept of territoriality. A territorial user interface is referred to as the set of interaction and physical surfaces, considered as parts or wholes, owned by a user involved in a dynamically-changing group collaboration in a given environment. For this purpose, we investigate five models covering the domain, the collaborative tasks, the users and the roles that play in the collaboration, the interaction surfaces involved in the collaboration, and the environment in which the collaboration takes places. For each model, intra-model relationships characterize static and dynamic relations. Across models, inter-model relationships dynamically map respective concepts.

## 9.2 International research visitors

### 9.2.1 Visits of international scientists

Andrei Ciortea, researcher at University of St. Gallen, visited Wimmics in September to work on RDF PubSub, security in CORESE and multi-agent systems on the Web.

## 9.3 European initiatives

### 9.3.1 FP7 & H2020 Projects

#### AI4EU

**Title:** A European AI On Demand Platform and Ecosystem

**Duration:** 2019 - 2021

**Coordinator:** THALES

**Partners:**

- AGENCIA ESTATAL CONSEJO SUPERIOR DE INVESTIGACIONES 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; [click here to know more](#)
- The implementation of industry-led pilots through the AI4EU platform, which demonstrates the capabilities of the platform to enable real applications and foster innovation; [click here to know more](#)
- 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; [click here to know more](#)
- 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.

## AI4Media

**Title:** AI4Media

**Duration:** 2020 - 2024

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

**Partners:**

<https://ai4media.eu/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.

### 9.3.2 Collaborations in European programs, except FP7 and H2020

#### 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: <http://hyperagents.gitlab.emse.fr/>

## 9.4 National initiatives

### PIA GDN ANSWER

**Participants** Fabien Gandon, Hai Huang, Vorakit Vorakitphan, Serena Villata, Elena Cabrio.

ANSWER stands for Advanced aNd Secured Web Experience and seaRch <sup>2</sup>. It is a GDN project (Grands Défis du Numérique) from the PIA program (Programme d'Investissements d'Avenir) on Big Data. The project is between four Inria research teams and the Qwant company.

The aim of the ANSWER project is to develop the new version of the Qwant <sup>3</sup> search engine by introducing radical innovations in terms of search criteria as well as indexed content and users' privacy.

The purpose is to strengthen everyone's confidence in the search engine and increase the effectiveness of Web search. Building trust in the search engine is based on innovations in (1) Security: computer security, privacy; (2) Completeness: completeness and heterogeneity of (re)sources; and (3) Neutrality: analysis, extraction, indexing, and classification of data.

Increasing the effectiveness of Web-based research relies on innovations related to (1) Relevance: variety and value of content taken into account, measurement of emotions carried by query results; (2) Interaction with the user: adaptation of the interfaces to the types of research; and (3) Performance: perceived relevance of results and response time.

The proposed innovations include:

- Design and develop models and tools for the detection of emotions in query results:
  - Ontology, thesaurus, linguistic resources
  - Metrics, indicators, classification of emotions
- Design and develop new crawling algorithms:
  - Dynamic crawling strategies
  - Crawlers and indexes for linked open data
- Ensure respect for privacy:
  - Detection of Internet tracking
  - Preventive display of tracing techniques
  - Certified security of automatic adaptation of ads to keywords entered by the user

## DGA CONFIRMA

**Participants** Elena Cabrio, Serena Villata.

This DGA project aims at automatically detecting fake news and limit their diffusion. In addition to identifying the communities propagating these fake news, we used methods from Natural Language Processing and Argumentation Theory to propose automatically extracted counter-arguments (adapted to target audience) from the existing reference press articles. These arguments allow to attack the false information detected in the fake news. Argument Mining techniques make it possible to (1) analyse the argumentation in natural language, for example by looking for the argumentative structures, identifying the relations of support or attack between the arguments; (2) locate the data related to specific information (related to fake news) on the Web. In the context of this project, Elena Cabrio and Serena Villata supervised the post-doc of Jerome Delobelle, now McF at University of Paris (LIPADE). Partners of the project: Storyzy, INRIA, Institut Jean Nicod. Duration of the project: 2018-2020.

## Ministry of Culture: MonaLIA 3.0

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

<sup>2</sup><https://project.inria.fr/answer/>

<sup>3</sup><http://www.qwant.com>

The objective of the MonaLIA project is to exploit the crossover of the automatic learning methods particularly applied to image analysis and knowledge-based representation and reasoning, in particular for the semantic indexing of annotated works and images in JocondeLab. The goal is to identify automated or semi-automatable tasks to improve the annotation. This project follows the preliminary project “MonaLIA 1” which established the state of the art in order to evaluate the potential and the combination of learning (notably deep learning) and the semantization of annotations on the case of JocondeLab. In the project MonaLIA we now want to go beyond the preliminary study and to design and build a prototype and the methods assisting the creation, the improvement and the maintenance of the metadata of the image database in order to assist the actors of the cultural world in their daily tasks. The preliminary study identified several possible coupling points between deep learning from non-necessarily structured data and reasoning from structured data. This project proposes to select the most promising of them to carry out a proof of concept combining these methods by focusing on the assistance to the annotation and curation tasks of the metadata of a real base to improve the contents, the course and exploitation thereafter.

#### 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.

Industrial transfer of some of the results of the WASABI project (partnership with AmpedStudio.com/Amp Track company) for integration of our software into theirs), SATT PACA.

Web site: <http://wasabihome.i3s.unice.fr>

#### ANR SIDES 3.0

**Participants** Catherine Faron, Olivier Corby, Antonia Ettore, Fabien Gandon, Alain Giboin, Mathis Le Quiniou, Franck Michel.

Partners: Université Grenoble Alpes, Inria, Ecole Normale Supérieure de Lyon, Viseo, Theia.

SIDES 3.0 is an ANR project which started in fall 2017. It is led by Université Grenoble Alpes (UGA) and its general objective is to introduce semantics within the existing SIDES educational platform<sup>4</sup> for medicine students, in order to provide them with added value educational services. Within this project Catherine Faron supervised the post-doctoral work of Oscar Rodriguez now research engineer in the TeachOnMars company, the master internship of Mathis Le Quiniou and is now supervising the PhD work of Antonia Ettore with Franck Michel. We are developing an approach to predict the success of students on training quizzes based on the knowledge graph representing their interactions with the pedagogical resources within the SIDES platform.

Web site: <https://www.uness.fr/anr/projets/dune/sides3.0>

#### ANR D2KAB

**Participants** Olivier Corby, Catherine Faron, Franck Michel.

---

<sup>4</sup><http://side-sante.org/>

Partners: LIRMM, INRA, 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: <http://www.d2kab.org>

### 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.

Web site: <https://dekalog.univ-nantes.fr/>

### Smart Enseigno

**Participants** Catherine Faron, Yuting Sun.

Partner: Educlever, Ludotic, Cabrilog, IFE

The Smart Enseigno project started in September 2019, led by Educlever. It is funded by the Ministry of National Education (MEN), within the Programme des Investissements d’Avenir (PIA2), action Partenariat d’innovation Intelligence artificielle (PI-IA) <sup>5 6</sup>. This project aims at developing resources and intelligent services within the Educlever platform for secondary school mathematics education. Within this project Catherine Faron supervised the work of Yuting Sun aiming to adapt the approach developed in the framework of the SIDES project to the Enseigno platform. This platform now relies on a knowledge graph to capture the interactions for students with pedagogical resources.

### DBpedia.fr

**Participants** Fabien Gandon, Franck Michel.

The DBpedia.fr project proposes the creation of a French chapter of the DBpedia database. This project was the first project of the Semanticpedia convention signed by the Ministry of Culture, the Wikimedia foundation and Inria.

Web site: <http://dbpedia.fr>

<sup>5</sup><https://eduscol.education.fr/pid29713/appels-a-projets-numeriques-des-investissements-d-avenir.html>

<sup>6</sup><https://primabord.eduscol.education.fr/P2IA>



## Convention between Inria and the Ministry of Culture

**Participants** 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 cross-road of the cultural domain and the digital sciences.

## Qwant-Inria Joint Laboratory

**Participants** Fabien Gandon.

We supervise the Qwant-Inria Joint Laboratory where joint teams are created and funded to contribute to the search engine research and development. The motto of the joint lab is Smart Search and Privacy with five research directions:

- Crawling, Indexing, Searching
- Execution platform, privacy by design, security, ethics
- Maps and navigation
- Augmented interaction, connected objects, chatbots, personal assistants
- Education technologies (EdTech)

We released the final, but confidential, report of the Qwant-Culture short-term project. This project aimed at identifying possibilities of exploiting the Qwant search engine to improve the search for information in the digital cultural resources of the French Ministry of Culture. Some possibilities have been selected to be the subject of research actions in the context a long-term project.

## CovidOnTheWeb - Covid Inria program

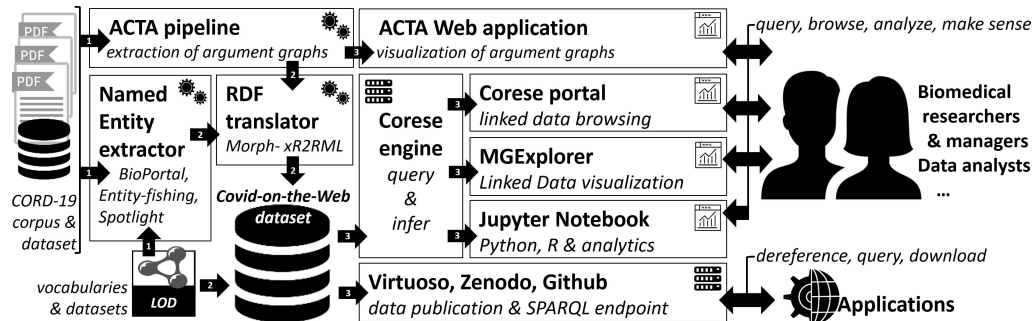
**Participants** Valentin Ah-Kane, Anna Bobasheva, Lucie Cadorel, Olivier Corby, Elena Cabrio, Jean-Marie Dormoy, Fabien Gandon, Raphaël Gazzotti, Alain Giboin, Abdelhadi Lebbar, Santiago Marro, Tobias Mayer, Aline Menin, Franck Michel, Andrea Tettamanzi, Serena Villata, Marco Winckler.

The project CovidOnTheWeb [48] 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 that extends and combines tools meant to process, analyze and enrich corpora such as the COVID-19 Open Research Dataset (CORD-19) that gathers 100,000+ full-text scientific articles related to the coronaviruses. The methods employed leverage knowledge representation, text mining, argument mining, as well as data visualization and exploration techniques.

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.

Among other tools, we rely on DBpedia Spotlight to identify and disambiguate NEs, and we used a local DBpedia instance to generate richer linksets linking NEs to other DBpedia chapters and Wikidata.

On top of this dataset, we have adapted Semantic Web tools (Corese, MGExplorer) to provide Linked Data visualizations that meet the expectations of the biomedical community. We are currently working on the implementation of data curation techniques that could be used to detect errors in the extraction and disambiguation of named entities. We plan for our future release of the dataset to use the latest English model of DBpedia Spotlight and then, in a next step, to detect entities in other languages with the same tool.



Web site: <https://github.com/Wimmics/CovidOnTheWeb>

## 9.5 Regional initiatives

### 3IA Côte d'Azur

**Participants** Catherine Faron, Fabien Gandon, Freddy Limpens, Andrea Tettamanzi, Serena Villata.

3IA Côte d'Azur is one of the four “Interdisciplinary Institutes of Artificial Intelligence”<sup>7</sup> 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 INRA. The project has also secured the support of more than 62 companies and start-ups.

We have three 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.

## 10 Dissemination

### 10.1 Promoting scientific activities

#### 10.1.1 Scientific events: organisation

##### General chair, scientific chair

- Marco Winckler was General Chair of the The 12th The ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS'2020), June 23-26, 2020, Sophia Antipolis, France and Associated Chair of the ACM Engineering Interactive Systems (EICS 2020), Sophia Antipolis, France.

##### Member of the organizing committees

- Marco Winckler was publicity chair of the International Conference on Web Engineering (ICWE'2020), Helsinki, Finland.

<sup>7</sup><https://3ia.univ-cotedazur.eu/>

### 10.1.2 Scientific events: selection

#### Chair of conference program committees

- Serena Villata was Program Chair of the 33rd International Conference on Legal Knowledge and Information Systems (JURIX-2020), Prague, Czech Republic, 9-11 December 2020 – virtual event due to COVID-19.
- Serena Villata was the Chair of the “Sister Conference Best Papers” track of the 29th International Joint Conference on Artificial Intelligence (IJCAI-2020).
- Elena Cabrio and Serena Villata were the Program Co-Chairs of the 7th Workshop on Argument Mining (ArgMining-2020) @COLING.
- Elena Cabrio was co-chair of the 4th Workshop on Natural Language for Artificial Intelligence @AIXIA conference.
- Marco Winckler was Technical Program co-Chair of the AVI 2020 - Advanced Visual Interfaces, September 28 - October 2, 2020 - Island of Ischia, Italy.

#### Member of the conference program committees

- Elena Cabrio was members of the Senior Program Committee of AAAI 2020 (Conference of the Association for the Advancement of Artificial Intelligence), ECAI 2020 (European Conference in Artificial Intelligence), and Program Committee members of EMNLP, COLING, ACL.
- Olivier Corby: European Semantic Web Conference ESWC, Graph Structures for Knowledge Representation and Reasoning GKR, International Conference on Knowledge Engineering and Knowledge Management EKAW, Ingénierie des Connaissances IC, International Joint Conference on Artificial Intelligence IJCAI, International Conference on Conceptual Structures ICCS, International Semantic Web Conference ISWC.
- Catherine Faron: Senior PC member of TheWebConf 2021 ; PC member of ESWC 2020 (European Semantic Web Conference), ISWC 2020 (Int. Semantic Web Conference), EKAW 2020 (Int. Conf. on Knowledge Engineering and Knowledge Management), Semantics 2020, ICCS 2020 (Int. Conference on Conceptual Structures), GKR 2020 (Int. workshop on Graph Structures for Knowledge Representation and Reasoning), IC 2020 (Ingénierie des Connaissances).
- Fabien Gandon: Senior PC (ACM International Conference on Information and Knowledge Management); PC ECAI2020 (European Conference in Artificial Intelligence); PC ESWC 2020 (European Semantic Web Conference); PC IJCAI-PRICAI 2020 (International Joint Conference on Artificial Intelligence) ; PC ISWC2020 (International Semantic Web Conference)
- Alain Giboin was PC member of IC 2020 (Ingénierie des Connaissances), VOILA 2020 (Visualization and Interaction for Ontologies and Linked Data)
- Serena Villata was member of the Senior Program Committee of AAAI 2020 (Conference of the Association for the Advancement of Artificial Intelligence), ECAI 2020 (European Conference in Artificial Intelligence), and Program Committee members of EMNLP, COLING, ACL, and was Area chair for “Sentiment Analysis, Stylistic Analysis, and Argument Mining” at ACL-2020.
- Franck Michel: PC member of the Int. Conference on Conceptual Structures (ICCS 2020), International Joint Conference on Artificial Intelligence (IJCAI-2020).
- Andrea Tettamanzi: Senior PC member of International Joint Conference on Artificial Intelligence (IJCAI-2020); PC member of AAAI 2021, CIKM 2020, ECAI2020 (European Conference in Artificial Intelligence), EKAW 2020 (Int. Conf. on Knowledge Engineering and Knowledge Management), ESWC 2020 (European Semantic Web Conference), EvoApplications (part of Evo\*) 2020, ICAART 2021, SUM 2020 (Scalable Uncertainty Management).

- Marco Winckler was member of the program committee of: the ACM SAC Track BPMM - Business Process Management & Modeling, Brno, Czech Republic. ; the ADVANCE 2020 workshop, Cancún, Mexico; the Brazilian Symposium on Human-Computer Interaction (IHC'2020), Diamantina, Brazil ; the HCSE 2020 8th International Conference on Human-Centered Software Engineering, Eindhoven, The Netherlands ; the IFIP IOT 2020 - 3rd IFIP International Internet of Things Conference, Amsterdam, The Netherlands ; the International Conference on Human-Computer Interaction - Interacción 2020, Malaga, Spain ; ; the International Conference on Web Engineering (ICWE'2020), Helsinki, Finland. ; the ManComp 2020: 5th Workshop on Managed Complexity, Riga, Latvia ; the MIDI2020 (Machine Intelligence Digital Interaction Conference); NORDICHI 2020, Nordic forum for Human-Computer Interaction (HCI), Estonia; the S-BPM ONE 2020, Bremen, Germany; SVR 2020 (Symposium on Virtual and Augmented Reality), Porto de Galinhas, Brazil; WEBIST 2020 – 17th International Conference on Web Information Systems and Technologies, Valletta, Malta.

### 10.1.3 Journal

#### Member of the editorial boards

- Catherine Faron: Editorial board member of Revue Ouverte d'Intelligence Artificielle <sup>8</sup>; Guest editor of the Semantic Web journal, Volume 12, Number 1 / 2021 (in press) <sup>9</sup>.
- Serena Villata was member of the Editorial Board of the journal "Artificial Intelligence and Law" <sup>10</sup>, of the journal "Argument and Computation" <sup>11</sup>, and of the journal "Journal of Web Semantics" <sup>12</sup>
- Marco Winckler became member of the editorial board of the Multimodal Technologies and Interaction – Open Access Journal (ISSN 2414-4088). <sup>13</sup>
- Marco Winckler become associated editor of the Journal Behaviour & Information Technology (Taylor & Francis).

#### Reviewing Activities

- Olivier Corby: Semantic Web Journal
- Catherine Faron, Michel Buffa: Journal of Web Semantics
- Andrea Tettamanzi: IEEE Access, Knowledge-Based Systems, Transactions of Fuzzy Systems.

### 10.1.4 Invited talks

- Serena Villata was invited speaker of the 3rd International Conference on Intelligent Technologies and Applications (INTAP-2020): "Artificial Machines Arguing For And With People". September 28-30, 2020, Gjøvik, Norvege <sup>14</sup>, and invited speaker of the Workshop on Dialogue, Explanation and Argumentation for Human-Agent Interaction, co-located with ECAI2020, September 7th, 2020 <sup>15</sup>.
- Elena Cabrio and Serena Villata were invited to present the Master Class organized by Telecom Valley: "Monitoring Cyberbullying through Message Classification and Social Network Analysis", November 2020, online <sup>16</sup>.
- Fabien Gandon was panelist of the ACM Web Science Conference 2020 Spotlight Panel 3 "Research Roadmap" <https://www.southampton.ac.uk/wsi/websci20-panels.page>
- Fabien Gandon gave an invited talk for ISWC 2020 Vision track <https://www.youtube.com/watch?v=b9GP00u2PTM>.

<sup>8</sup><https://roia.centre-mersenne.org/>

<sup>9</sup><http://www.semantic-web-journal.net/>

<sup>10</sup><https://www.springer.com/journal/10506/editors>

<sup>11</sup><https://www.iospress.nl/journal/argument-computation/>

<sup>12</sup><https://www.journals.elsevier.com/journal-of-web-semantics/editorial-board>

<sup>13</sup><https://www.mdpi.com/journal/mti>

<sup>14</sup><http://jdconline.net/intap/keynotes.html>

<sup>15</sup><https://sites.google.com/view/dexahai-at-ecai2020/home>

<sup>16</sup><http://www.telecom-valley.fr/masterclass-sophia-2020/>

### 10.1.5 Leadership within the scientific community

Fabien Gandon is a member of 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 and steering committee of the ISWC conference.

Marco Winckler is Secretary of the IFIP TC13 on Human-Computer Interaction.

### 10.1.6 Scientific expertise

- Fabien Gandon : ERC-StG 2020 reviewer.
- Catherine Faron : member of the ANR scientific evaluation committee "Artificial Intelligence" (CE23) ; member of the scientific evaluation committee "National Research Data Infrastructure" (NFDI) of the German research agency (DFG) ; reviewer of project proposals for the ANR regional call for projects Résilience Grand Est ; reviewer for the National Research Programme "Covid-19" of the Swiss National Science Foundation (SNSF) ; reviewer for the SESAME call for projects of Région Ile de France ; scientific referent of the Inria Learning Lab.
- Andrea Tettamanzi: reviewer of a CIFRE thesis proposal for ANRT; reviewer for the Swiss National Science Foundation.
- Marco Winckler: CHIST-ERA & ERC-AdG reviewer.

### 10.1.7 Research administration

- Andrea Tettamanzi and Marco Winckler are responsible for the SPARKS team of I3S.
- Fabien Gandon : evaluation committee for 3IA Côte d'Azur chairs ; Vice-director of Research Inria Sophia Antipolis ; jury DR2 Inria ; jury PEDR Inria ; Evaluation Committee of Inria.
- Catherine Faron : member of the HCERES committee in charge of the evaluation of the LIRIS laboratory ; General Treasurer of the French Society for Artificial Intelligence (AFIA) ; member of the steering committee of the AFIA college on Knowledge Engineering ; member of the 2020 evaluation committee of Inria ; member of the CPRH 27 commission at Université Côte d'Azur.

## 10.2 Teaching Supervision Juries

### 10.2.1 Teaching

- Licence: Andrea Tettamanzi, Introduction à l'Intelligence Artificielle, 27 h ETD, L2, UCA, France.
- Licence: Michel Buffa, JavaScript, 40h, L3 MIAGE - Univ Côte d'Azur, France.
- Licence: Elena Cabrio, Web Technologies, 80 hours, (Portail Sciences de la Vie), UCA, France.
- Licence: Elena Cabrio, Internship supervision, 27 hours, (L3MIAGE), UCA, France.
- Master: Michel Buffa, Web technologies front and back end, 40h, M1, UCA, France.
- Master: Michel Buffa, Server Side JavaScript and modern front-end frameworks, 60h, M2 MIAGE - Univ Côte d'Azur NTDP and MBDS, UCA, France.
- Master: Michel Buffa, Introduction to AI, MIAGE - Univ Côte d'Azur Master 1 and Master 2 IA2.
- Master: Michel Buffa, Multiplayer game programming, IA for games.
- Master: Elena Cabrio, Computational Linguistics, 30 hours, (Lettres), UCA, France.
- Master: Elena Cabrio, Natural Language Processing for AI, 30 hours, (M1 INFO), UCA, France.
- Master and Licence: Elena Cabrio, Responsible of the internship programme, 40 hours, (L3 and M2 MIAGE), UCA, France.

- Master: Olivier Corby, Semantic Web, 20h, Polytech Nice Sophia - Univ Côte d'Azur, France.
- Master: Catherine Faron, Web languages, 48h, M1, Polytech Nice Sophia - Univ Côte d'Azur, UCA.
- Master: Catherine Faron, Semantic Web technologies (EN), 48h, M2 Informatique, Polytech Nice Sophia - Univ Côte d'Azur, UCA.
- Master: Catherine Faron, Knowledge Engineering 28h, M2 Informatique, Polytech Nice Sophia - Univ Côte d'Azur, UCA.
- Master: Catherine Faron, Semantic Web technologies (EN), 30h, M1 Data Science, UCA.
- Master: Catherine Faron, XML technologies, 16h, M2 IMAFA, Polytech Nice Sophia - Univ Côte d'Azur, UCA.
- Master: Catherine Faron, Projects and Internship tutoring, 32h, M2, Polytech Nice Sophia - Univ Côte d'Azur, UCA.
- Master: Fabien Gandon, Integrating Semantic Web technologies in Data Science developments, 78 h, M2, DSTI, France.
- Master: Oscar Rodríguez Rocha, Web of Data, 15h, M2, Polytech Nice Sophia - Univ Côte d'Azur, France.
- Master: Oscar Rodríguez Rocha, Knowledge Engineering, 10h, M2, Polytech Nice Sophia - Univ Côte d'Azur, France.
- Master: Andrea Tettamanzi, Logic for AI, 30 h ETD, M1, UCA, France.
- Master: Andrea Tettamanzi, Web, 30 h ETD, M1, UCA, France.
- Master: Andrea Tettamanzi, Algorithmes Évolutionnaires, 24.5 h ETD, M2, UCA, France.
- Master: Andrea Tettamanzi, Modélisation del l'Incertitude, 24.5 h ETD, M2, UCA, France.
- Licence (L3/SI3): Marco Winckler, Introduction to Human-Compute Interaction. 40 h ETD, Polytech Nice Sophia - Univ Côte d'Azur, France.
- Master (M2/SI5): Marco Winckler, Design adn Evaluation of Interactive Systems. 40 h ETD, Polytech Nice, France.
- Master (M2/SI5): Marco Winckler, Interaciton Techniques. 10 h ETD, Polytech Nice Sophia - Univ Côte d'Azur, France.
- Master (M2 DS4H): Accessibilité et Design Universel. 15 h ETD, UCA, France.
- Master (M2/SI5): Introduction to Scientific Research. 6 h EDT, Polytech Nice Sophia - Univ Côte d'Azur, France.
- Master (M2): Introduction to Scientific Research. 6 h EDT, UCA, France.
- Master (M2 MBDS): Visualization de données. 15 h EDT, UCA, France.
- Master (M1 SDAI): Visualization de données. 15 h EDT, UCA, France.
- Master 2: coordinator of the 5th year UE TER (Travaux de Recherche et Etude). 15h EDT, Polytech Nice Sophia - Univ Côte d'Azur, France.

### **E-learning**

- Mooc: Michel Buffa, "JavaScript Intro" published first in Juin 2017 on the EDx platform (MIT/Harvard), still active and updated regularly.

- Mooc: Michel Buffa, "HTML5 Coding Essentials and Best Practices"
- Mooc: Michel Buffa, "HTML5 Apps and Games", also on EDx, are still active and updated regularly. More than 700.000 registered users since 2015 for these MOOCS.
- Mooc: Fabien Gandon, Olivier Corby & Catherine Faron, Web of Data and Semantic Web (FR), 7 weeks, <http://www.france-universite-numerique.fr/>, Inria, France Université Numérique, self-paced course 41002, Education for Adults, 10324 learners registered for 2020, <https://www.fun-mooc.fr/courses/course-v1:inria+41002+self-paced/about>
- Mooc: Fabien Gandon, Olivier Corby & Catherine Faron, Introduction to a Web of Linked Data (EN), 4 weeks, <http://www.france-universite-numerique.fr/>, Inria, France Université Numérique, self-paced course 41013, Education for Adults, 3827 learners registered for 2020, <https://www.fun-mooc.fr/courses/course-v1:inria+41013+self-paced/about>
- Mooc: Fabien Gandon, Olivier Corby & Catherine Faron, Web of Data (EN), 4 weeks, <https://www.coursera.org/>, Coursera, self-paced course Education for Adults, 3228 learners registered, <https://coursera.org/learn/web-data>

### 10.2.2 Supervision

- PhD in progress: **Maroua Tikat**, Interactive multimedia visualization for the exploration of multidimensional metadata database of popular music, UCA, Michel Buffa, Marco Winckler.
- PhD in progress: **Shihong Ren**, Which tools for music composition and real-time signal processing on the web, in a collaborative approach? UCA, Michel Buffa, Université de St Etienne, Laurent Pottier.
- PhD in progress: **Molka Dhouib**, *Knowledge engineering in the sourcing domain for the recommendation of providers*, UCA, Catherine Faron, Andrea Tettamanzi.
- PhD in progress: **Ahmed El Amine Djebri**, *Uncertainty in Linked Data*, UCA, Andrea Tettamanzi, Fabien Gandon.
- PhD in progress: **Antonia Ettore**, *Artificial Intelligence for Education and Training: Knowledge Representation and Reasoning for the development of intelligent services in pedagogical environments*, UCA, Catherine Faron, Franck Michel.
- PhD: **Michael Fell**, *Natural Language Processing of Song Lyrics*, UCA, Co-supervision Elena Cabrio and Fabien Gandon, July 2020. [3]
- PhD: **Raphaël Gazzotti**, *Knowledge graphs based extension of patients' files to predict hospitalization*, UCA, Catherine Faron, Fabien Gandon, April [59].
- PhD in progress: **Santiago Marro**, *Argument-based Explanatory Dialogues for Medicine*, UCA 3IA, Elena Cabrio and Serena Villata.
- PhD in progress: **Nicholas Halliwell**, *Explainable and Interpretable Prediction*, UCA, Fabien Gandon.
- PhD: **Tobias Mayer**, *Argument Mining for Clinical Trials*, UCA, Serena Villata, Elena Cabrio and Céline Poudat (UCA), December 2020 [5].
- PhD in progress: **Thu Huong Nguyen**, *Mining the Semantic Web for OWL Axioms*, Andrea Tettamanzi, UCA.
- PhD in progress: **Mahamadou Toure**, *Models and architectures for restricted and local mobile access to the Data Web*, UCA, Fabien Gandon, Moussa Lo (UGB, Senegal).
- PhD in progress: **Vorakit Vorakitphan**, *Argumentation and Emotions Emotion Detection with Adaptive Sentiment Analysis*, Elena Cabrio, Serena Villata, UCA.

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

### Internship

- Master internship: **ElMahdi Ammari**, GUI builder for WebAudio plugins (WebComponents) developed as part of the WASABI project. Integration into the FAUST IDE.
- Master Internship: **Valeria Bellusci**, Evolutionary Axiom Discovery from Populated Knowledge Bases, UCA, Andrea Tettamanzi.
- Master internship: **Matthis Lequinou**, *Prediction of student's success on the TeachOnMars Knowledge Graph*, UCA, Catherine Faron & Oscar Rodríguez Rocha.
- Master internship: **Zineb Rahhali**, Machine learning to associate songs with presets of instruments and audio effects encoded in WebAudio.
- Master internship: **Yuting Sun**, *Prediction of student's success on the Educlever Knowledge Graph*, UCA, Catherine Faron & Franck Michel.
- Master internship: **Abdelhadi Lebbar**, *Exploitation de données géospatiales à l'intersection entre graphes de connaissance et données d'imagerie satellitaire*, Franck Michel & Marco Winckler.
- Master apprenticeship: **Benjamin Molinet**, Enriching the WASABI semantic dataset with NLP and audio processing.
- Master 2 internship: **Valentin Ah-Kane**. *LinkedDataVis-bis - Vers un modèle de transformation générique pour la visualisation interactive de données linked-data*, UCA.
- Master 1 intership: **Jean-Marie Dormoy**. *Adaptation de l'outil de visualisation LinkedDataViz au domaine du COVID-19 : Interrogation et visualisation de données liées*. UCA, Alain Giboin & Olivier Corby

### 10.2.3 Juries

Michel Buffa: Reviewer of Pasquale LISENA PhD : “Recommandation musicale basée sur la connaissance : modèles, algorithmes et recherche exploratoire”, defended October 11th, 2019, EURECOM – Sophia Antipolis

Elena Cabrio:

- Reviewer of the PhD committee of Giovanni Siragusa, University of Turin (Italy), 2020.
- Member of the PhD committee of Gabriel Meseguer Brocal, Ircam, 2020.

Catherine Faron:

- reviewer of Pierre Larmande's HDR, entitled *Intégration de Données Multi-Echelles et Extraction de Connaissances en Agronomie: Exemples et Perspectives*, defended on September 11 at Université de Montpellier;
- member of Konstantin Todorov's HDR, entitled *Towards a Web of Structured Knowledge: Methods, Applications and Perspectives*, defended on June 29 at Université de Montpellier;
- member of Patricia Serrano Alvarado's HDR, entitled *Protecting user data in distributed systems*, defended on June 16 at Université de Nantes;



- reviewer of Yves Mercadier's PhD thesis, entitled *Classification automatique de textes par réseaux de neurones profonds : application au domaine de la santé*, defended on November 17 at Université de Montpellier;
- member of Pierre-Henri Paris' PhD thesis jury, entitled *Identity in RDF Knowledge Graphs*, defended on June 17 at Sorbonne Université;
- external member of the monitoring committee of Stella Zevio's PhD thesis at Université Paris Nord;
- external member of the monitoring committee of Francesco Bariatti's PhD thesis at Université de Rennes;
- external member of the monitoring committee of Charbel Obeid's PhD thesis at Université de Lyon.
- member of the monitoring committee of Thu Huong Nguyen's PhD thesis at Université Côte d'Azur.

Fabien Gandon:

- external member of the monitoring committee of Hicham Hossayni PhD thesis at Telecom SudParis, Institut Polytechnique de Paris;
- reviewer of Thomas Minier PhD thesis, entitled *Web Preemption for Querying the Linked Open Data*, defended on November 10th, 2020 at Université de Nantes, France;
- reviewer of Pierre Monnin PhD thesis, entitled *Matching and mining in knowledge graphs of the Web of data Applications in pharmacogenomics*, defended on December 16th, 2020 at Université de Lorraine, Loria, France;
- president of Elena Cabrio HDR thesis, entitled *Artificial Intelligence to Extract, Analyze and Generate Knowledge and Arguments from Texts to Support Informed Interaction and Decision Making* defended 22/10/2020, Université Côte d'Azur.
- reviewer for the Fondazione Bruno Kessler (FBK) Tenure Track program.

Alain Giboin :

- Invited Member of the PhD thesis jury of Marie Destandau (thesis title: "Path-Based Interactive Visual Exploration of Knowledge Graphs"), December 18, Paris-Saclay University.

Serena Villata:

- Gia-Lac Tran, EURECOM. Title of the thesis: "Advances of Deep Gaussian Processes: Calibration and Sparsification". Role: member of the jury. PhD defense: 2020.
- Benjamin Moreau, University of Nantes. Title of the thesis: "Facilitating Reuse on the Web of Data", Role: reviewer. PhD defense: 2020.

Andrea Tettamanzi:

- Reviewer of Victor Eduardo Fuentes, PhD, Université du Québec à Montréal, *Méta alignement méta heuristique*, 6 octobre 2020;
- PhD Committee Chair for Edson Florez, *Adverse drug reactions detection in clinical notes*, Université Côte d'Azur, 01/07/2020;
- PhD Committee Chair for Raphaël Gazzotti, *Prédiction d'hospitalisation par la génération de caractéristiques extraites de graphes de connaissances*, Université Côte d'Azur, 30/04/2020;
- PhD Committee Chair for Gérald Rocher, *Évaluation de l'Effectivité des Systèmes Ambiants*, Université Côte d'Azur, 10/02/2020;

Marco Winckler:

- Reviewer of Jérôme Dupire HDR. “Vers une Accessibilité Accessible”. Presented on December 4th 2020, Université Paris 8 Vincennes Saint Denis, Paris, France.
- Reviewer of Tanguy Giuffrida PhD. “Fuzzy4U : un système d’adaptation des IHM en logique floue pour l’accessibilité”. Presented on December 12th 2020, Université Grenoble Alpes, Grenoble, France.
- Jury member of Aline Menin PhD. “eSTIME: a visualization framework for assisting a multi-perspective analysis of daily mobility data”. Presented on November 26th 2020, Université Grenoble Alpes, Grenoble, France.

#### 10.2.4 Teaching Administration

- Michel Buffa: director of MIAGE - Univ Côte d’Azur.
- Elena Cabrio: vice director of MIAGE - Univ Côte d’Azur.
- Catherine Faron: coordinator of the Web and AI option of the 5th year of Polytech Nice Sophia - Univ Côte d’Azur engineering school; pedagogical responsible of continuous training for the computer science department of Polytech Nice Sophia - Univ Côte d’Azur.
- Marco Winckler: coordinator of the Human-Computer Interaction track of the 5th year of Polytech Nice Sophia - Univ Côte d’Azur engineering school.

### 10.3 Popularization

#### 10.3.1 Articles and contents

Fabien Gandon:

- Article in “Annales des Mines Enjeux Numériques” – about “Une toile de fond pour le Web : lier les données et lier leurs vocabulaires sur la toile, pour un Web plus accessible aux machines” [12].
- Contributor to book / whitepaper “Éducation et numérique, Défis et enjeux” [69].
- Contributor again of the second version of the book / whitepaper “Artificial Intelligence: Current challenges and Inria’s engagement” [65].
- Animation of reading sessions at the Knowledge Graph Conference book club on chapters of the textbook “Semantic Web for the Working Ontologist” [51]<sup>17</sup>.

Elena Cabrio and Fabien Gandon are two characters in the comic book “Les défis de l’intelligence artificielle – Un reporter dans les labos de recherche” [66].

Publication of the third edition of the textbook “Semantic Web for the Working Ontologist” [51] with Fabien Gandon as new co-author.

## 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](https://doi.org/10.1145/3382097). URL: <https://hal.inria.fr/hal-02939606>.
- [2] 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>.

<sup>17</sup>[https://www.knowledgegraph.tech/events\\_espresso/book-club-sem-web-working-ontologist/](https://www.knowledgegraph.tech/events_espresso/book-club-sem-web-working-ontologist/)

- [3] 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>.
- [4] 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>.
- [5] T. Mayer. ‘Argument Mining on Clinical Trials’. Theses. Université Côte d’Azur, Dec. 2020. URL: <https://hal.archives-ouvertes.fr/tel-03137784>.

## 11.2 Publications of the year

### International journals

- [6] S. Basso, A. Ceselli and A. G. B. Tettamanzi. ‘Random sampling and machine learning to understand good decompositions’. In: *Annals of Operations Research* 284.2 (2020), pp. 501–526. DOI: [10.1007/s10479-018-3067-9](https://doi.org/10.1007/s10479-018-3067-9). URL: <https://hal.inria.fr/hal-02319521>.
- [7] B. Berendt, F. Gandon, S. Halford, W. Hall, J. Hendler, K. E. Kinder-Kurlanda, E. Ntoutsis and S. Staab. ‘Web Futures: Inclusive, Intelligent, Sustainable The 2020 Manifesto for Web Science’. In: *Dagstuhl Manifestos* (2021). DOI: [10.4230/DagMan.9.1.1](https://doi.org/10.4230/DagMan.9.1.1). URL: <https://hal.inria.fr/hal-03189474>.
- [8] J. Bowen, J. Vanderdonck and M. Winckler. ‘A Glimpse into the Past, Present, and Future of Engineering Interactive Computing Systems’. In: *Proceedings of the ACM on Human-Computer Interaction*. Proceedings of the ACM on Human-Computer Interaction (HCI) 2020 4 : EICS.article 71 (18th June 2020), pp. 1–32. DOI: [10.1145/3394973](https://doi.org/10.1145/3394973). URL: <https://hal.archives-ouvertes.fr/hal-03138759>.
- [9] A. Ciorrea, S. Mayer, S. Bienz, F. Gandon and O. Corby. ‘Autonomous search in a social and ubiquitous Web’. In: *Personal and Ubiquitous Computing* (9th June 2020). DOI: [10.1007/s00779-020-01415-1](https://doi.org/10.1007/s00779-020-01415-1). URL: <https://hal.inria.fr/hal-03135088>.
- [10] M. Corazza, S. Menini, E. Cabrio, S. Tonelli and S. Villata. ‘A Multilingual Evaluation for Online Hate Speech Detection’. In: *ACM Transactions on Internet Technology* 20.2 (25th May 2020), pp. 1–22. DOI: [10.1145/3377323](https://doi.org/10.1145/3377323). URL: <https://hal.archives-ouvertes.fr/hal-02972184>.
- [11] R. David, L. Mabile, A. Specht, S. Stryeck, M. Thomsen, M. Yahia, C. Jonquet, L. Dollé, D. Jacob, D. Bailo, E. Bravo, S. Gachet, H. Gunderman, J.-E. Hollebecq, V. Ioannidis, Y. Le Bras, E. Lerigoleur and A. Cambon-Thomsen. ‘FAIRness Literacy: The Achilles’ Heel of Applying FAIR Principles’. In: *CODATA Data Science Journal* 19.32 (11th Aug. 2020), pp. 1–11. DOI: [10.5334/dsj-2020-032](https://doi.org/10.5334/dsj-2020-032). URL: <https://hal.archives-ouvertes.fr/hal-02483307>.
- [12] F. Gandon. ‘Une toile de fond pour le Web : lier les données et lier leurs vocabulaires sur la toile, pour un Web plus accessible aux machines’. In: *Annales des Mines - Enjeux Numériques* (10th June 2020). URL: <https://hal.inria.fr/hal-02748219>.
- [13] V. Leone, L. Di Caro and S. Villata. ‘Taking stock of legal ontologies: a feature-based comparative analysis’. In: *Artificial Intelligence and Law* 28.2 (2020), pp. 207–235. DOI: [10.1007/s10506-019-09252-1](https://doi.org/10.1007/s10506-019-09252-1). URL: <https://hal.archives-ouvertes.fr/hal-03065217>.
- [14] S. Ren, S. Letz, Y. Orlarey, R. Michon, D. Fober, M. Buffa and J. Lebrun. ‘Using Faust DSL to Develop Custom, Sample Accurate DSP Code and Audio Plugins for the Web Browser’. In: *Journal of the Audio Engineering Society* 68.10 (30th Nov. 2020). URL: <https://hal.inria.fr/hal-03087763>.
- [15] T. Rocha Silva, M. Winckler and H. Trættemberg. ‘Ensuring the Consistency between User Requirements and Task Models: A Behavior-Based Automated Approach’. In: *Proceedings of the ACM on Human-Computer Interaction*. Proceedings of the ACM on Human-Computer Interaction (HCI) 4 : EICS.article 77 (18th June 2020), pp. 1–32. DOI: [10.1145/3394979](https://doi.org/10.1145/3394979). URL: <https://hal.archives-ouvertes.fr/hal-03138755>.
- [16] G. Rocher, J.-Y. Tigli, S. Lavirotte and N. Le Thanh. ‘Effectiveness assessment of Cyber-Physical Systems’. In: *International Journal of Approximate Reasoning* 118 (2020), pp. 112–132. DOI: [10.1016/j.ijar.2019.12.002](https://doi.org/10.1016/j.ijar.2019.12.002). URL: <https://hal.archives-ouvertes.fr/hal-02441654>.

- [17] O. Rodríguez Rocha, C. Faron Zucker, A. Giboin and A. Lagarrigue. ‘Automatic Generation of Questions from DBpedia’. In: *International Journal of Continuing Engineering Education and Life-Long Learning* x.1 (2020), p. 1. DOI: [10.1504/IJCEELL.2020.10024221](https://doi.org/10.1504/IJCEELL.2020.10024221). URL: <https://hal.archives-ouvertes.fr/hal-02571170>.

#### International peer-reviewed conferences

- [18] N. Broders, C. Martinie, P. Palanque, M. Winckler and K. Halunen. ‘A Generic Multimodels-Based Approach for the Analysis of Usability and Security of Authentication Mechanisms’. In: *HCSE 2020: Human-Centered Software Engineering*. HCSE 2020 - 8th International Conference on Human-Centered Software Engineering - IFIP WG 13.2 International Working Conference. Vol. 12481. Lecture Notes in Computer Science book series (LNCS). Eindhoven/ Online, Netherlands: Springer, 25th Nov. 2020, pp. 61–83. DOI: [10.1007/978-3-030-64266-2\\_4](https://doi.org/10.1007/978-3-030-64266-2_4). URL: <https://hal.archives-ouvertes.fr/hal-03079818>.
- [19] M. Buffa and J. Lebrun. ‘A FAUST-based re-creation of the power amp stage for WebAudio-based simulations of guitar tube amplifiers’. In: *IFC 2020 - Second International Functional Audio Stream (Faust) Conference*. Saint-Denis / Virtual, France, 1st Dec. 2020. URL: <https://hal.inria.fr/hal-03087768>.
- [20] M. Buffa, J. Lebrun, S. Ren, S. Letz, Y. Orlarey, R. Michon and D. Fober. ‘Emerging W3C APIs opened up commercial opportunities for computer music applications’. In: *The Web Conference 2020 - DevTrack*. Taipei, Taiwan, 20th Apr. 2020. URL: <https://hal.inria.fr/hal-02557901>.
- [21] L. Cadorel and A. G. B. Tettamanzi. ‘Mining RDF Data of COVID-19 Scientific Literature for Interesting Association Rules’. In: *WI-IAT’20 - IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*. The 2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT’20), 14-17 December 2020, a fully virtual conference. Melbourne, Australia, 2020. URL: <https://hal.inria.fr/hal-03084029>.
- [22] O. Cocarascu, E. Cabrio, S. Villata and F. Toni. ‘Dataset Independent Baselines for Relation Prediction in Argument Mining’. In: *COMMA 2020 - 8th International Conference on Computational Models of Argument*. Vol. 326. Frontiers in Artificial Intelligence and Applications. Perugia, Italy, 8th Sept. 2020, pp. 45–52. DOI: [10.3233/FAIA200490](https://doi.org/10.3233/FAIA200490). URL: <https://hal.archives-ouvertes.fr/hal-02972180>.
- [23] A.-M. Déry-Pinna, P. R. Gonin, M. Winckler, C. Kolski, S. Lepreux and J. Vanderdonckt. ‘Entrain, exploring new territorial user interfaces’. In: *EICS ’20 Companion: Companion Proceedings of the 12th ACM SIGCHI Symposium on Engineering Interactive Computing Systems*. EICS 2020 - 12th ACM SIGCHI Symposium on Engineering Interactive Computing Systems. Article 7. Sophia Antipolis, France: ACM: Association for Computing Machinery, 23rd June 2020, pp. 1–4. DOI: [10.1145/3393672.3398648](https://doi.org/10.1145/3393672.3398648). URL: <https://hal.archives-ouvertes.fr/hal-03138763>.
- [24] A. Ettore, O. R. Rocha, C. Faron Zucker, F. Michel and F. Gandon. ‘A Knowledge Graph Enhanced Learner Model to Predict Outcomes to Questions in the Medical Field’. In: *EKAW 2020 - 22nd International Conference on Knowledge Engineering and Knowledge Management*. Bolzano, Italy, 16th Sept. 2020. URL: <https://hal.inria.fr/hal-02915414>.
- [25] R. Gazzotti, C. Faron Zucker, F. Gandon, V. Lacroix-Hugues and D. Darmon. ‘Injection of Automatically Selected DBpedia Subjects in Electronic Medical Records to boost Hospitalization Prediction’. In: *SAC 2020 - 35th ACM/SIGAPP Symposium On Applied Computing*. Brno, Czech Republic, 30th Mar. 2020. DOI: [10.1145/3341105.3373932](https://doi.org/10.1145/3341105.3373932). URL: <https://hal.archives-ouvertes.fr/hal-02389918>.
- [26] H. Huang and F. Gandon. ‘Learning URI Selection Criteria to Improve the Crawling of Linked Open Data (Extended Abstract)’. In: *The 29th International Joint Conference on Artificial Intelligence (IJCAI 2020)*. Yokohama, Japan, 7th Jan. 2021. URL: <https://hal.inria.fr/hal-03064912>.

- [27] A. Laadhar, E. Abrahão and C. Jonquet. ‘Analysis of Term Reuse, Term Overlap and Extracted Mappings across AgroPortal Semantic Resources’. In: EKAU 2020 - 22nd International Conference on Knowledge Engineering and Knowledge Management. Bozen-Bolzano, Italy, Sept. 2020. URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-02945172>.
- [28] A. Laadhar, E. Abrahão and C. Jonquet. ‘Investigating One Million XRefs in Thirthy Ontologies from the OBO World’. In: ICBO’20 - 11th International Conference on Biomedical Ontologies. Vol. 2807. G.1-12. Bozen-Bolzano, Italy: CEUR Workshop Proceedings, Sept. 2020. URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-02945170>.
- [29] D. Malchiodi, C. Da Costa Pereira and A. G. B. Tettamanzi. ‘Classifying Candidate Axioms via Dimensionality Reduction Techniques’. In: *Modeling Decisions for Artificial Intelligence - 17th International Conference, MDAI 2020, Sant Cugat, Spain, September 2-4, 2020, Proceedings; Modeling Decisions for Artificial Intelligence - 17th International Conference, MDAI 2020, Sant Cugat, Spain, September 2-4, 2020, Proceedings*. MDAI 2020 - 17th International Conference on Modeling Decisions for Artificial Intelligence. Sant Cugat, Spain: Springer, 26th Aug. 2020, pp. 179–191. DOI: [10.1007/978-3-030-57524-3\\_15](https://doi.org/10.1007/978-3-030-57524-3_15). URL: <https://hal.inria.fr/hal-02931396>.
- [30] F. Michel, G. Olivier, B. Ledentec and B. Community. ‘Unleash the Potential of your Website! 180,000 webpages from the French Natural History Museum marked up with Bioschemas/Schema.org biodiversity types’. In: TDWG 2020 annual conference. Vol. 4. Biodiversity Information Science and Standards (BISS). Virtual, France, 29th Sept. 2020. DOI: [10.3897/biss.4.59046](https://doi.org/10.3897/biss.4.59046). URL: <https://hal.archives-ouvertes.fr/hal-02976710>.
- [31] T. H. Nguyen and A. G. B. Tettamanzi. ‘A Multi-Objective Evolutionary Approach to Class Disjointness Axiom Discovery’. In: WI-IAT 2020 - IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology. Melbourne/ Virtual, Australia, 14th Dec. 2020. URL: <https://hal.inria.fr/hal-03082579>.
- [32] T. H. Nguyen and A. G. B. Tettamanzi. ‘Grammatical Evolution to Mine OWL Disjointness Axioms Involving Complex Concept Expressions’. In: CEC 2020 - IEEE Congress on Evolutionary Computation. Glasgow, United Kingdom: IEEE, 19th July 2020, pp. 1–8. DOI: [10.1109/CEC48606.2020.9185681](https://doi.org/10.1109/CEC48606.2020.9185681). URL: <https://hal.inria.fr/hal-02936253>.
- [33] T. H. Nguyen and A. G. B. Tettamanzi. ‘Using Grammar-Based Genetic Programming for Mining Disjointness Axioms Involving Complex Class Expressions’. In: *Proceedings of the 25th International Conference on Conceptual Structures (ICCS 2020)*. ICCS 2020 - 25th International Conference on Conceptual Structures. Vol. 12277. Lecture Notes in Computer Science. Bozen-Bolzano, Italy: Springer, 10th Sept. 2020, pp. 18–32. DOI: [10.1007/978-3-030-57855-8\\_2](https://doi.org/10.1007/978-3-030-57855-8_2). URL: <https://hal.inria.fr/hal-02936193>.
- [34] C. Pasquier, C. Da Costa Pereira and A. G. B. Tettamanzi. ‘Extending a Fuzzy Polarity Propagation Method for Multi-Domain Sentiment Analysis with Word Embedding and POS Tagging’. In: *Frontiers in Artificial Intelligence and Applications*. ECAI 2020 - 24th European Conference on Artificial Intelligence. Vol. 325. Santiago de Compostela, Spain: IOS Press, 29th Aug. 2020, pp. 2140–2147. DOI: [10.3233/FAIA200338](https://doi.org/10.3233/FAIA200338). URL: <https://hal.inria.fr/hal-02936130>.
- [35] S. Ren, L. Pottier and M. Buffa. ‘From Diagram to Code: a Web-based Interactive Graph Editor for Faust DSP Design and Code Generation’. In: IFC 2020 - Second International Functional Audio Stream (Faust) Conference. Saint-Denis / Virtual, France, 1st Dec. 2020. URL: <https://hal.inria.fr/hal-03087778>.
- [36] M. Tounsi Dhoub, C. Faron and A. Tettamanzi. ‘Injection of Knowledge in a Sourcing Recommender System’. In: WI-IAT’20 - IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology. Melbourne / Virtual, Australia, 14th Dec. 2020. URL: <https://hal.inria.fr/hal-02996442>.
- [37] M. Toure, K. Guidedi, F. Gandon, M. Lo and C. Guéret. ‘MoRAI: Geographic and Semantic Overlay Network for Linked Data Access with Intermittent Internet Connectivity’. In: WI-IAT 2020 - IEEE/WIC/ACM International Joint Conference On Web Intelligence And Intelligent Agent Technology (WI-IAT ’20). Melbourne, Australia, 2021. URL: <https://hal.inria.fr/hal-03099724>.

- [38] V. Vorakitphan, M. Guerini, E. Cabrio and S. Villata. ‘Regrexit or not Regrexit: Aspect-based Sentiment Analysis in Polarized Contexts’. In: COLING’2020 - 28th International Conference on Computational Linguistics (COLING’2020). Barcelona / Online, Spain, 8th Dec. 2020. URL: <https://hal.inria.fr/hal-03031812>.

#### National peer-reviewed Conferences

- [39] H. Daoud, M. Tounsi Dhouib, J. Rancati, C. Faron and A. Tettamanzi. ‘A Hybrid Bi-LSTM-CRF Model for Sequence Labeling Applied to the Sourcing Domain’. In: PFIA-APIA 2020 - 5ème Conférence Nationale sur les Applications Pratiques de l’Intelligence Artificielle. Angers, France, 29th June 2020. URL: <https://hal.inria.fr/hal-02932095>.

#### Conferences without proceedings

- [40] M. Corazza, S. Menini, E. Cabrio, S. Tonelli and S. Villata. ‘Hybrid Emoji-Based Masked Language Models for Zero-Shot Abusive Language Detection’. In: EMNLP 2020 - Conference on Empirical Methods in Natural Language Processing. Virtual, France, 16th Nov. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02972203>.
- [41] J. Delobelle, A. Delamaire, E. Cabrio, R. Ruti and S. Villata. ‘Sifting the Arguments in Fake News to Boost a Disinformation Analysis Tool’. In: NL4AI 2020 - 4th Workshop on Natural Language for Artificial Intelligence. Online, Italy, 25th Nov. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02990781>.
- [42] A. E. A. Djebri, A. Ettore and J. Mortara. ‘Towards a Linked Open Code’. In: ESWC 2021 - The 18th Extended Semantic Web Conference. Heraklion / Virtual, Greece, 6th June 2021. URL: <https://hal.archives-ouvertes.fr/hal-03190617>.
- [43] A. E. A. Djebri, A. G. B. Tettamanzi and F. Gandon. ‘Task-Oriented Uncertainty Evaluation for Linked Data Based on Graph Interlinks’. In: EKAW 2020 - 22nd International Conference on Knowledge Engineering and Knowledge Management. Bozen-Bolzano, Italy, 16th Sept. 2020. DOI: 10.1007/978-3-030-61244-3\_15. URL: <https://hal.archives-ouvertes.fr/hal-02933190>.
- [44] M. Fell, E. Cabrio, E. Korfed, M. Buffa and F. Gandon. ‘Love Me, Love Me, Say (and Write!) that You Love Me: Enriching the WASABI Song Corpus with Lyrics Annotations’. In: LREC 2020 - 12th edition of the Language Resources and Evaluation Conference. Marseille, France, 11th May 2020. URL: <https://hal.archives-ouvertes.fr/hal-03133188>.
- [45] T. Mayer. ‘Enriching Language Models with Semantics’. In: ECAI 2020 - 24th European Conference on Artificial Intelligence. Santiago de Compostela / Online, Spain, 29th Aug. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02879286>.
- [46] T. Mayer, E. Cabrio and S. Villata. ‘Transformer-based Argument Mining for Healthcare Applications’. In: *Proceedings of ECAI 2020 - 24th European Conference on Artificial Intelligence*. ECAI 2020 - 24th European Conference on Artificial Intelligence. Santiago de Compostela / Online, Spain, 29th Aug. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02879293>.
- [47] T. Mayer, S. Marro, E. Cabrio and S. Villata. ‘Generating Adversarial Examples for Topic-dependent Argument Classification’. In: COMMA 2020 - 8th International Conference on Computational Models of Argument. Perugia, Italy, 8th Sept. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02933266>.
- [48] 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>.

- [49] M. Romero, N. Timus, S. Raisin, I. Mirbel and I. Aliferis. ‘Accompagner la transformation des pratiques enseignantes à l’Université Côte d’Azur par le biais d’une approche centrée sur la professionnalisation des enseignant.e.s’. In: AIPU 2020 - 31e Congrès de l’Association Internationale de Pédagogie Universitaire. Québec, Canada, 2020. DOI: [10.1111/cogs.12626](https://hal.archives-ouvertes.fr/hal-02552836). URL: <https://hal.archives-ouvertes.fr/hal-02552836>.

### Scientific books

- [50] J. Abdelnour Nocera, A. Parmaxi, M. Winckler, F. Loizides, C. Ardito, G. Bhutkar and P. Dannenmann. *Beyond Interactions: INTERACT 2019 IFIP TC 13 Workshops 2019, Revised Selected Papers*. Vol. 11930. Lecture Notes in Computer Science book series (LNCS). Paphos, Cyprus: Springer, Jan. 2020. DOI: [10.1007/978-3-030-46540-7](https://hal.archives-ouvertes.fr/hal-03138795). URL: <https://hal.archives-ouvertes.fr/hal-03138795>.
- [51] D. Allemang, J. Hendler and F. Gandon. *Semantic Web for the Working Ontologist*. 3. ACM, 15th June 2020. DOI: [10.1145/3382097](https://hal.inria.fr/hal-02939606). URL: <https://hal.inria.fr/hal-02939606>.

### Scientific book chapters

- [52] F. Bertonecello, M.-J. Ouriachi, C. Da Costa Pereira, A. G. B. Tettamanzi, L. Purdue and R. Ajroud. ‘Using ABM to explore the role of socio-environmental interactions on Ancient Settlement Dynamics.’ In: *Human history and digital future, Proceeding of the 46th Computer Applications and Quantitative Methods in Archaeology international conference (CAA 2018), Tübingen, mars 2018*. 2021. URL: <https://halshs.archives-ouvertes.fr/halshs-03098473>.
- [53] A. Reuter, K. Boudaoud, M. Winckler, A. Abdelmaksoud and W. Lemrazzeq. ‘Secure Email - A Usability Study’. In: *FC 2020: Workshops, International Conference on Financial Cryptography and Data Security*. Vol. 12063. Lecture Notes in Computer Science book series (LNCS, volume 12063). 7th Aug. 2020, pp. 36–46. DOI: [10.1007/978-3-030-54455-3\\_3](https://hal.archives-ouvertes.fr/hal-03138769). URL: <https://hal.archives-ouvertes.fr/hal-03138769>.
- [54] A. G. B. Tettamanzi, D. Emsellem, C. Da Costa Pereira, A. Venerandi and G. Fusco. ‘Possibilistic Estimation of Distributions to Leverage Sparse Data in Machine Learning’. In: *Information Processing and Management of Uncertainty in Knowledge-Based Systems. IPMU 2020. Information Processing and Management of Uncertainty in Knowledge-Based Systems - 18th International Conference, IPMU 2020, Lisbon, Portugal, June 15-19, 2020, Proceedings, Part I*. Lisbon, Portugal: Springer, 5th June 2020, pp. 431–444. DOI: [10.1007/978-3-030-50146-4\\_32](https://hal.archives-ouvertes.fr/hal-02885825). URL: <https://hal.archives-ouvertes.fr/hal-02885825>.

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

- [55] J. Bowen, J. Vanderdonckt and M. Winckler, eds. *EICS '20 Companion: Companion Proceedings of the 12th ACM SIGCHI Symposium on Engineering Interactive Computing Systems*. 12th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS 2020). Sophia Antipolis (virtual), France: ACM : Association for Computing Machinery, New York, NY, United States, June 2020. DOI: [10.1145/3393672](https://hal.archives-ouvertes.fr/hal-03138784). URL: <https://hal.archives-ouvertes.fr/hal-03138784>.
- [56] G. Tortora, G. Vitiello and M. Winckler, eds. *AVI '20: Proceedings of the International Conference on Advanced Visual Interfaces*. International Conference on Advanced Visual Interfaces (AVI 2020). Ischia Island, Italy: ACM : Association for Computing Machinery, New York, NY, United States, Sept. 2020. DOI: [10.1145/3399715](https://hal.archives-ouvertes.fr/hal-03138779). URL: <https://hal.archives-ouvertes.fr/hal-03138779>.

### Doctoral dissertations and habilitation theses

- [57] E. Cabrio. ‘Artificial Intelligence to Extract, Analyze and Generate Knowledge and Arguments from Texts to Support Informed Interaction and Decision Making’. Université Côte d’Azur, 22nd Oct. 2020. URL: <https://hal.inria.fr/tel-03084380>.
- [58] M. Fell. ‘Natural language processing for music information retrieval : deep analysis of lyrics structure and content’. Université Côte d’Azur, 12th May 2020. URL: <https://tel.archives-ouvertes.fr/tel-02587910>.

- [59] R. Gazzotti. ‘Knowledge graphs based extension of patients’ files to predict hospitalization’. Université Côte d’Azur, 30th Apr. 2020. URL: <https://tel.archives-ouvertes.fr/tel-03135236>.
- [60] T. Mayer. ‘Argument Mining on Clinical Trials’. Université Côte d’Azur, 17th Dec. 2020. URL: <https://tel.archives-ouvertes.fr/tel-03209489>.
- [61] A. Menin. ‘eSTIME: a visualization framework for assisting a multi-perspective analysis of daily mobility data’. Grenoble UGA - Université Grenoble Alpes, 26th Nov. 2020. URL: <https://hal.archives-ouvertes.fr/tel-03081054>.

### Reports & preprints

- [62] O. Cocarascu, E. Cabrio, S. Villata and F. Toni. *A Dataset Independent Set of Baselines for Relation Prediction in Argument Mining*. 15th Dec. 2020. URL: <https://hal.archives-ouvertes.fr/hal-03067403>.
- [63] M. Destandau, O. Corby, J.-D. Fekete and A. Giboin. *Path Outlines: Browsing Path-Based Summaries of Linked Datasets*. 25th Feb. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02490218>.

## 11.3 Other

### Scientific popularization

- [64] F. Gandon. *Les IA comprennent-elles ce qu’elles font ?* 20th Nov. 2020. URL: <https://hal.inria.fr/hal-03135131>.

## 11.4 Cited publications

- [65] B. Braunschweig. *Artificial Intelligence: Current challenges and Inria’s engagement - Inria white paper*. Livre blanc Inria. Aug. 2016. URL: <https://hal.inria.fr/hal-01564589>.
- [66] J. Dres. *Les défis de l’intelligence artificielle : un reporter dans les labos de recherche*. Paris: First, 2021.
- [67] F. Gandon. ‘The three ‘W’ of the World Wide Web call for the three ‘M’ of a Massively Multidisciplinary Methodology’. In: *10th International Conference, WEBIST 2014*. Ed. by V. Monfort and K.-H. Krempels. Vol. 226. Web Information Systems and Technologies. Barcelona, Spain: Springer International Publishing, Apr. 2014. DOI: [10.1007/978-3-319-27030-2](https://doi.org/10.1007/978-3-319-27030-2). URL: <https://hal.inria.fr/hal-01223236>.
- [68] 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: *5th 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>.
- [69] G. Giraudon, P. Guitton, M. Romero, D. Roy and T. Viéville. *Éducation et numérique, Défis et enjeux*. Livre Blanc Inria. Inria, Dec. 2020, p. 137. URL: <https://hal.inria.fr/hal-03051329>.
- [70] F. Michel, S. Terceirie, A. Ettorre, O. Gargominy and C. Faron Zucker. ‘Assisting Biologists in Editing Taxonomic Information by Confronting Multiple Data Sources using Linked Data Standards’. In: *Biodiversity Next*. Vol. 3. Biodiversity Information Science and Standards 37421. Leiden, Netherlands, Oct. 2019. DOI: [10.3897/biss.3.37421](https://doi.org/10.3897/biss.3.37421). URL: <https://hal.archives-ouvertes.fr/hal-02168164>.
- [71] M. A. Winckler, R. Cava, E. Barboni, P. Palanque and C. Freitas. ‘Usability aspects of the inside-in approach for ancillary search tasks on the web’. In: *15th Human-Computer Interaction (INTERACT)*. Ed. by J. Abascal, S. Barbosa, M. Fetter, T. Gross, P. Palanque and M. Winckler. Vol. LNCS-9297. Human-Computer Interaction – INTERACT 2015 Part II. Bamberg, Germany: Springer, Sept. 2015, pp. 211–230. DOI: [10.1007/978-3-319-22668-2\\_18](https://doi.org/10.1007/978-3-319-22668-2_18). URL: <https://hal.archives-ouvertes.fr/hal-01343048>.