



SEMAPRO 2022

The Sixteenth International Conference on Advances in Semantic Processing

ISBN: 978-1-61208-991-1

November 13 - 17, 2022

Valencia, Spain

SEMAPRO 2022 Editors

Jose Miguel Jimenez Herranz, Universitat Politecnica de Valencia, Spain

SEMAPRO 2022

Forward

The Sixteenth International Conference on Advances in Semantic Processing (SEMAPRO 2022), held between November 13 and November 17, 2022, continued a series of events that were initiated considering the complexity of understanding and processing information. Semantic processing considers contextual dependencies and adds to the individually acquired knowledge emergent properties and understanding. Hardware and software support and platforms were developed for semantically enhanced information retrieval and interpretation. Searching for video, voice and speech [VVS] raises additional problems to specialized engines with respect to text search. Contextual searching and special patterns-based techniques are current solutions.

With the progress on ontology, web services, semantic social media, semantic web, deep web search /deep semantic web/, semantic deep web, semantic networking and semantic reasoning, SEMAPRO 2022 constituted the stage for the state-of-the-art on the most recent advances.

We take here the opportunity to warmly thank all the members of the SEMAPRO 2022 technical program committee, as well as all the reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors that dedicated much of their time and effort to contribute to SEMAPRO 2022. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

We also gratefully thank the members of the SEMAPRO 2022 organizing committee for their help in handling the logistics and for their work that made this professional meeting a success.

We hope that SEMAPRO 2022 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in the field of semantic processing. We hope that Valencia provided a pleasant environment during the conference and everyone saved some time to enjoy the charm of the city.

SEMAPRO 2022 Chairs

SEMAPRO 2022 Steering Committee

Fabio Grandi, University of Bologna, Italy

Sandra Lovrenčić, University of Zagreb, Croatia

Michele Melchiori, Università degli Studi di Brescia, Italy

Wladyslaw Homenda, Warsaw University of Technology, Poland

Tim vor der Brück, Lucerne University of Applied Sciences and Arts, Switzerland

Els Lefever, LT3 | Ghent University, Belgium

Sofia Athenikos, Twitter, USA

SEMAPRO 2022 Publicity Chair

Mar Parra, Universitat Politecnica de Valencia, Spain

Sandra Viciano Tudela, Universitat Politecnica de Valencia, Spain

SEMAPRO 2022

COMMITTEE

SEMAPRO 2022 Steering Committee

Sandra Lovrenčić, University of Zagreb, Croatia
Michele Melchiori, Università degli Studi di Brescia, Italy
Fabio Grandi, University of Bologna, Italy
Sofia Athenikos, Twitter, USA
Wladyslaw Homenda, Warsaw University of Technology, Poland
Tim vor der Brück, Lucerne University of Applied Sciences and Arts, Switzerland
Els Lefever, LT3 | Ghent University, Belgium

SEMAPRO 2022 Publicity Chair

Mar Parra, Universitat Politecnica de Valencia, Spain
Sandra Viciano Tudela, Universitat Politecnica de Valencia, Spain

SEMAPRO 2022 Technical Program Committee

Witold Abramowicz, Poznan University of Economics, Poland
Harry Agius, Brunel University London, UK
Jose María Alvarez Rodríguez, Carlos III University of Madrid, Spain
Abdel-Karim Al-Tamimi, Higher Colleges of Technology, UAE
Sofia Athenikos, Twitter, USA
Arvind Bansal, Kent State University, USA
Giuseppe Berio, Université de Bretagne Sud | IRISA, France
Floris Bex, Utrecht University & University of Tilburg, Netherlands
Loris Bozzato, Fondazione Bruno Kessler, Trento, Italy
Zouhaier Brahmia, University of Sfax, Tunisia
Okan Bursa, Ege University, Turkey
Ozgu Can, Ege University, Turkey
Rodrigo Capobianco Guido, São Paulo State University (UNESP), Brazil
Damir Cavar, Indiana University, USA
Alberto Cetoli, QBE Europe, UK
David Chaves-Fraga, Universidad Politécnica de Madrid, Spain
Ioannis Chrysakis, FORTH-ICS, Greece / Ghent University, Belgium
Cláudio de Souza Baptista, Federal University of Campina Grande, Brazil
Christian Dirschl, Wolters Kluwer Deutschland GmbH, Germany
Milan Dojchinovski, InfAI | Leipzig University, Germany / Czech Technical University in Prague, Czech Republic
Julio Cesar Duarte, Instituto Militar de Engenharia, Rio de Janeiro, Brazil
Enrico Francesconi, IGSG - CNR, Italy
Rolf Fricke, Condat AG, Berlin, Germany
Panorea Gaitanou, Greek Ministry of Justice, Athens, Greece
Bilel Gargouri, MIRACL Laboratory | University of Sfax, Tunisia

Fabio Grandi, University of Bologna, Italy
Damien Graux, ADAPT Centre - Trinity College Dublin, Ireland
Jingzhi Guo, University of Macau, Macau SAR, China
Bidyut Gupta, Southern Illinois University Carbondale, USA
Prabhakar Gupta, Parth Group, India
Shun Hattori, The University of Shiga Prefecture, Japan
Tobias Hellmund, Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB, Germany
Tracy Holloway King, Amazon, USA
Timo Homburg, Mainz University of Applied Sciences, Germany
Wladyslaw Homenda, Warsaw University of Technology, Poland
Tzung-Pei Hong, National University of Kaohsiung, Taiwan
Helmut Horacek, DFKI/Saarland University, Germany
Thomas Hubauer, Siemens AG Corporate Technology, Germany
Sergio Ilarri, University of Zaragoza, Spain
Agnieszka Jastrzebska, Warsaw University of Technology, Poland
Naouel Karam, Fraunhofer FOKUS, Berlin, Germany
Armita Khajeh Nassiri, Paris Saclay University | LISN | CNRS, France
Young-Gab Kim, Sejong University, Korea
Stasinios Konstantopoulos, Institute of Informatics and Telecommunications, NCSR "Demokritos", Greece
Petr Kremen, Czech Technical University in Prague, Czech Republic
Jaroslav Kuchař, Czech Technical University in Prague, Czech Republic
Chun-Ming Lai, Tunghai University, Taiwan
Kyu-Chul Lee, Chungnam National University, South Korea
Els Lefever, LT3 | Ghent University, Belgium
Antoni Ligeża, AGH-UST Kraków, Poland
Usha Lokala, University of South Carolina, USA
Giuseppe Loseto, Polytechnic University of Bari, Italy
Sandra Lovrenčić, University of Zagreb, Croatia
Federica Mandreoli, Università di Modena e Reggio Emilia, Italy
Miguel A. Martínez-Prieto, University of Valladolid, Segovia, Spain
Miguel Felix Mata Rivera, UPIITA-IPN, Mexico
Michele Melchiori, Università degli Studi di Brescia, Italy
Dimitri Metaxas, Rutgers University, USA
Mohamed Wiem Mkaouer, Rochester Institute of Technology, USA
Luis Morgado da Costa, Nanyang Technological University, Singapore
Fadi Muheidat, California State University San Bernardino, USA
Yotaro Nakayama, Technology Research & Innovation Nihon Unisys, Ltd., Tokyo, Japan
Nikolay Nikolov, SINTEF Digital, Norway
Fabrizio Orlandi, ADAPT Centre | Trinity College Dublin, Ireland
Peera Pacharintanakul, TOT, Thailand
Peteris Paikens, University of Latvia - Faculty of Computing, Latvia
Panagiotis Papadakos, FORTH-ICS | University of Crete, Greece
Silvia Piccini, Institute Of Computational Linguistics "A. Zampolli" (CNR-Pisa), Italy
Vitor Pinheiro de Almeida, Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Brazil
Livia Predoiu, Otto-von-Guericke-Universität Magdeburg, Germany
Matthew Purver, Queen Mary University of London, UK
Francisco José Quesada Real, Universidad de Cádiz, Spain

Irene Renau, Pontificia Universidad Católica de Valparaíso, Colombia
Tarmo Robal, Tallinn University of Technology, Estonia
Christophe Roche, University Savoie Mont-Blanc, France
Dmitri Roussinov, University of Strathclyde, UK
Michele Ruta, Politecnico di Bari, Italy
Minoru Sasaki, Ibaraki University, Japan
Fabio M. A. Santos, Northern Arizona University, USA
Lenhart Schubert, University of Rochester, USA
Wieland Schwinger, Johannes Kepler University Linz (JKU) | Inst. f. Telekooperation (TK), Linz, Austria
Floriano Scioscia, Polytechnic University of Bari, Italy
Carlos Seror, Independent Researcher, Spain
Saeedeh Shekarpour, University of Dayton, USA
Liana Stanescu, University of Craiova, Romania
Mark Steedman, University of Edinburgh, Scotland, UK
George Tambouratzis, Institute for Language and Speech Processing, Athena R.C., Greece
Christos Tryfonopoulos, University of the Peloponnese, Greece
L. Alfonso Ureña-López, Universidad de Jaén, Spain
Taketoshi Ushiyama, Kyushu University, Japan
Sirje Virkus, Tallinn University, Estonia
Daiva Vitkute-Adzgauskienė, Vytautas Magnus University, Lithuania
Tim vor der Brück, Lucerne University of Applied Sciences and Arts, Switzerland
Heba Wageeh, British University in Egypt, Cairo, Egypt
Rita Zaharah Wan-Chik, Universiti Kuala Lumpur, Malaysia
Xiaofan Wang, Institute of Software | Chinese Academy of Sciences, China
Wai Lok Woo, Northumbria University, UK
Congyu "Peter" Wu, University of Texas at Austin, USA
Qiong Wu, AT&T Labs Research, USA
Roberto Yus, University of California, Irvine, USA
Stefan Zander, University of Applied Sciences Darmstadt, Germany
Martin Zelm, INTEROP-VLabBrussels, Belgium
Chao Zhang, University of Fukui, Japan
Shuai Zhao, New Jersey Institute of Technology, USA
Lu Zhou, Kansas State University, USA
Qiang Zhu, University of Michigan - Dearborn, USA

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission to reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

Representing Online Debates in the Context of E-Journalism <i>Elisjana Ymeralli, Giorgos Flouris, Vasilis Eftymiou, Katerina Papantoniou, Theodore Patkos, Georgios Petasis, Nikiforos Pittaras, Giannis Roussakis, and Elias Tzortzakakis</i>	1
Towards Exchanging Wearable-PGHD with EHRs: Developing a Standardized Information Model for Wearable-Based Patient Generated Health Data <i>Abdullahi Abubakar Kawu, Dympna O'Sullivan, and Lucy Hederman</i>	8

Representing Online Debates in the Context of E-Journalism

Elisjana Ymeralli*, Giorgos Flouris*, Vasilis Efthymiou*, Katerina Papantoniou*, Theodore Patkos*,
Georgios Petasis†, Nikiforos Pittaras†, Giannis Roussakis* and Elias Tzortzakakis*

*Institute of Computer Science, FORTH
Heraklion, Greece

Emails: {ymeralli, fgeo, vefthym, papanton, patkos, rousakis, tzortzak}@ics.forth.gr

†Institute of Informatics and Telecommunications, NCSR "Demokritos"
Athens, Greece

Emails: {petasis, pittarasnikif}@iit.demokritos.gr

Abstract—An important need of journalists writing opinion articles is the ability to identify, obtain and understand online arguments and opinions covering various perspectives of the debate at hand, a task that cannot be fully accomplished using simple keyword search. This paper describes research on analyzing Greek news articles from a variety of news sources, which considers the internal structure of arguments, in addition to their textual content. We describe a suite of tools for mining, representing and reasoning with real arguments, using semantic technologies, and argue that our tools can enhance future newsroom processes in the domain of online journalism.

Keywords—online journalism; computational argumentation; argument mining; reasoning.

I. INTRODUCTION

Quality journalism is inherently dependent on reliable and well-justified information delivery, raising the need for automated tools to support journalists in the era of digital information and rapid spread of news feed. One of the crucial components of quality journalism is the ability to identify, search and navigate efficiently in existing journalistic articles to find important arguments related to a topic of interest.

Although standard keyword search is of undeniable value in identifying relevant articles, one cannot use simple keywords to search on the basis of the structure of an argument, or on the basis of its relationships to other arguments. To support these needs, the DebateLab project [1] aims to develop a suite of tools that will allow the journalist, or the interested citizen, to navigate in journalistic articles and understand better their argumentative structure, and, eventually, the main points of both sides of important public debates. Our current focus lies on news articles and sources in the Greek language. However, our approach can easily be extended to other languages.

In this paper, we focus on the underlying representation structure for arguments, which was developed in order to support the above applications. In particular, we will present an ontological model (called *Onto4JARGs - Ontology for Journalistic Arguments* [2]) that we developed to represent, store and reason with arguments, articles and their constituents, in the context of e-journalism, as well as the process that

leads to the identification and ingestion of this information in Onto4JARGs.

In the following, we start by presenting the relevant background in the literature (Section II). Section III introduces the main components of Onto4JARGs, used for representing, storing and reasoning with arguments found in journalistic articles. Section IV outlines the pipeline of the ingestion process for identifying and transforming data into ontological information, whereas Section V presents the user-related concepts of Onto4JARGs. We conclude in Section VI.

II. BACKGROUND

A. Related Work

An increasing number of journalistic platforms exist in the literature for harvesting news-related content from the Web, and for analysing and further enriching data with relevant information from various Knowledge Bases (such as Wikipedia). These platforms often use knowledge graphs [3] and other semantic techniques [4] [5] that automatically analyse and enrich news material, and leverage theories and techniques from the field of artificial intelligence [6] and natural language processing [7] [8] to identify, classify, and process news events in a more meaningful way [9].

An example of such a work is described in [10], which describes a prototype for harvesting news-related content and social media messages, using an ontology for representing news items semantically, and well-defined methods from the field of artificial intelligence. Similarly, the *Neptuno* [11] describes an ontology that is useful for semantic search and browsing capabilities, as well as for visualizing content. *NewsReader* [12] [13] is a tool for analysing web news texts semantically and enriching them with relevant information from reference Knowledge Bases in order to build an event-centric knowledge graph. Further, the *EventRegistry* [14] news platform is used to extract data from RSS feeds and link them with relevant information about locations, people and organizations in real time.

Although the aforementioned systems provide useful informative services for journalists, our research is unique in combining many of these features and enriching them

with additional functionalities, such as the integration with state-of-the-art ontologies (AIF [15] – Subsection II-B), the identification of argument relationships using computational argumentation methods (Subsection IV-B) and the quantitative evaluation of arguments using sophisticated scoring algorithms (Subsection V-A).

B. Argument Interchange Format (AIF)

The AIF ontology [15] is an abstract ontology for representing argumentation information and relationships among arguments. Its aim is to serve as a “blueprint” towards the definition of more specific and application-dependent ontologies for argumentative information. It is thus closer to a high-level, generic conceptual model for argumentation. Here, we describe the specification used by the Argumentation Research Group at the University of Dundee [16], which is available in various formats [17].

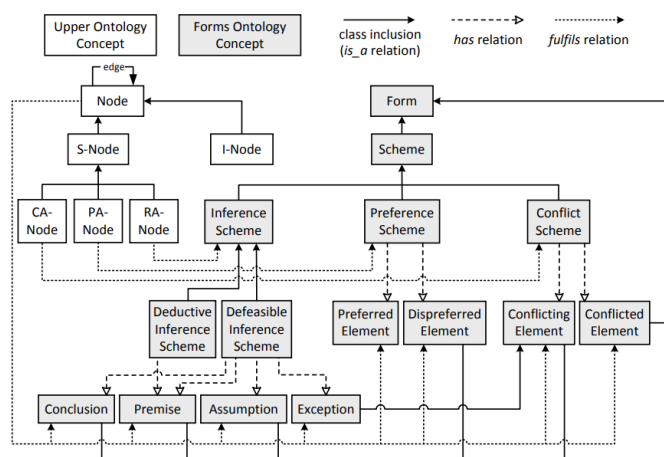


Figure 1. The AIF specification [15].

The main classes of the AIF ontology are shown in Figure 1. AIF consists of two components, namely the Upper Ontology (white boxes) and the Forms Ontology (gray boxes).

The Upper Ontology describes “nodes”, which essentially represent arguments and their components. There are two main types of nodes, namely I-Nodes and S-Nodes. Information nodes (I-Node) are used to represent the content of arguments and represent claims and premises that depend on the domain of discourse. Schema nodes (S-Node) represent applications of schemes, i.e., independent patterns of argumentative reasoning. There are three types of S-Nodes, namely CA-Nodes, PA-Nodes and RA-Nodes. CA-Nodes (Conflict Application Nodes) represent conflicts among other nodes, PA-Nodes (Preference Application Nodes) represent preferences, whereas RA-Nodes (Rule of Inference Application Nodes) represent the application of some inference scheme to develop arguments.

The Forms Ontology defines the types of statements and schemes typically used in argumentation. It contains several classes which embody the general principles for actually capturing the pattern of reasoning, which can be an inference rule

(*Inference Scheme*), a conflicting rule (*Conflict Scheme*), and a preference rule (*Preference Scheme*). Hence, the individual RA-, CA-, and PA-nodes fulfil these schemes to define whether a conclusion (say c) is inferred, or attacked from the premises (say p_i), as well as denote preferences over nodes.

C. Computational Argumentation

Computational argumentation is the field of study which deals with the representational and reasoning aspects that determine how arguments and argumentative processes can be represented in a computer system, and how the outcome of an argumentative process can be automatically determined [18]. Work in computational argumentation is often classified as either *structured* or *abstract* argumentation.

Structured argumentation concerns itself mainly with the internal structure of an argument, how it should be represented, and how this internal structure determines the relationships between arguments [19] for the representation of domain knowledge. Typically, an argument consists of a set of premises (say Δ) and a conclusion (say c) such that $\Delta \vdash c$, where the \vdash relationship corresponds to the inference relation of the underlying logic. Thus, the argument structure can be viewed as a sequence of statements (i.e., premises) which are often expressed in favor of or against other statements (i.e., given as conclusion), relative to a knowledge graph of structured relations and arguments.

All relations between arguments (e.g., attack, support) are determined by viewing the logical relationships among the argumentative units of the respective arguments, including conclusions and/or premises, as they are understood from the given text (i.e., news article) of the argument mining output. For further explanations on how two or more arguments are related with each other, see also Figure 3 and the related analysis in Subsection III-B.

On the other hand, abstract argumentation ignores the internal structure of arguments and considers only their relationships (e.g., attack, support), attempting to determine the semantics (i.e., the acceptable arguments) given a set of arguments and their relationships. This strand of work was initiated by the work of Dung [20] who viewed a debate as a directed graph, whose nodes are the arguments and the arrows represent attack relationships among them. Since the work of Dung [20], numerous semantics and extensions of the above simple framework have been defined.

III. MAIN ARGUMENTATIVE CONCEPTS IN ONTO4JARGS

For the purposes of DebateLab, we developed *Onto4JARGs* [2], a *Resource Description Framework* (RDF) [21] - based argumentation ontology that fits the needs of the project. *Onto4JARGs* is heavily based on AIF [15] [16], and essentially enhances and extends its abstract model (Figure 1), reusing most of its main concepts.

Figure 2 shows the main classes and properties of the ontology [2]. In Figure 2, dark blue circles represent classes from the AIF specification, while light blue circles are new classes, introduced by us to represent concepts necessary for

arguments and statements from different documents are identified (Subsection IV-E). We analyze these steps in more details in the sequel.

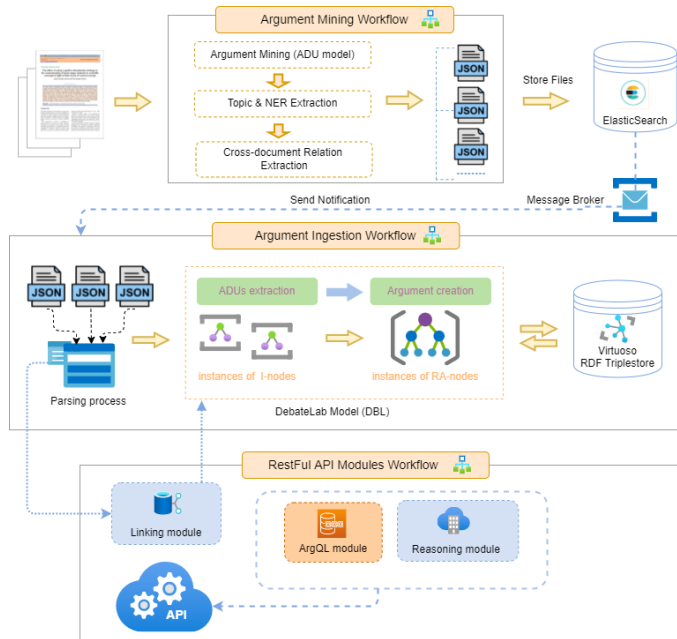


Figure 4. The ingestion workflow of DebateLab [1].

A. News articles detection

The main source of information in DebateLab is journalistic articles crawled from the web, represented through the *Article* class. An article consists of ADUs, organised into arguments, as explained in Subsection III-A. Articles and its constituents (ADUs, arguments) are detected by the argument mining process [25] [26], which is a Natural Language Processing task, aiming to detect and identify argumentative structures from text using Machine Learning methods.

In particular, we utilize transformer models to embed text [27], namely the popular *Bidirectional Encoder Representations from Transformers* (BERT) architecture [28], which defines pre-training schemes on natural language understanding (NLU) tasks such as masked language modeling and next sentence prediction. Due to limited resources in our target language and domain, we take advantage pretrained models [29] in Greek in order to leverage generic knowledge on language structure, prior to fine-tuning on a manually annotated dataset of 150 news articles crawled from the web.

Using these labelled data, we build ADU detection, relation and stance classification models via training token and sequence classification heads on the transformer via supervised fine-tuning. These classifiers yield ADU segments and relation / stance labels on ADU pairs, reaching macro-F1 scores of 0.56, 0.93 and 0.89 respectively in a 3-fold cross-validation setting. Note that, the ADU detector was evaluated using sequence-level rather than token-level matching of predicted and ground truth token sequences. This measure is far harder to satisfy compared to token-level evaluation.

Corresponding baseline macro-F1 scores from uniform random predictions are 0.22, 0.26 and 0.39, indicating that all components perform far above naive baselines and are able to extract argumentative structure from the documents. There is room for improvement in ADU detection, which showcases reduced performance compared to the REL and STANCE components. We estimate that this occurs due to the limited training dataset, label imbalance in the training data as well as the sequence-oriented evaluation. Performance is thus expected to improve for all components as more documents are annotated and added to the training pool.

The process takes as input a set of Greek documents (articles), crawled from the Web, and analyzes them in order to deliver the required information in a structured manner, through appropriate JSON files, which contain the following information regarding an article:

- A list of ADUs classified as premises, claims or major claims.
- The association of each premise to a single claim, and the association of each claim to a major claim. Two types of associations are identified, depending on whether the premise/claim attempts to validate or refute the claim/major claim (denoted by *sup*/*att* respectively).
- A list of metadata (i.e., *identifier*, *content*, *confidence score*) connected to each ADU node.
- A set of topics, tags and categories associated with the article.

B. Creating structured arguments

The next step in the ingestion pipeline is to process the raw information provided by argument mining in order to generate arguments and their relationships. The approach we follow for generating structured arguments identifies three different cases (see Figure 5, and the description below), depending on the relationships among the premises and their respective conclusion, as identified by argument mining. Note that for this particular part of the analysis we do not differentiate between claims and major claims, i.e., the argument's conclusion ADU can be either a claim or a major claim.

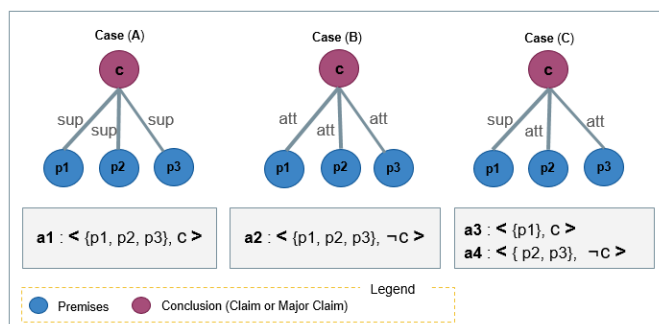


Figure 5. The argument-generation algorithm [2].

Case (A): If all the ADUs (say p_i) associated to a claim or major claim (say c) are associated with it through a *sup* relation, then we create a single argument, whose conclusion

is c and whose premises are all the individual p_i . This case is visualised in the left-most part of Figure 5.

Case (B): If all the ADUs (say p_i) associated to a claim or major claim (say c) are associated with it through an *att* relation, then the process is more complex. First, we create a new, artificial ADU, denoted by $\neg c$, which represents the negation of c . Then, we create a single argument, whose conclusion is $\neg c$ and whose premises are all the individual p_i . This corresponds to the second case of Figure 5. Note that the reason for creating the artificial ADU $\neg c$ is related to the requirement of structured argumentation that the premises of an argument imply the argument’s conclusion [19].

Case (C): If a claim or major claim c is associated to different ADUs using both the *att* and *sup* relations, then we apply the above two cases for each group separately. In particular, we generate one argument whose conclusion is c and whose premises are the ADUs associated with c using the *sup* relation (as per Case (A)), and one argument whose conclusion is $\neg c$ and whose premises are the ADUs associated with c using the *att* relation (as per Case (B)). This case is visualised in the right-most part of Figure 5.

C. Enrichment of argumentative data

The *stance* of a UAI represents its attitude (*for* or *against*) towards the topics of an article. To determine the stance, we leverage argument associations. In particular, the ADUs and their associations (*sup/att*) that were identified by the argument mining procedure (Subsection IV-A) in any given article, create a hierarchical structure, whose nodes are the ADUs that appear in the article, and these nodes are connected with the respective *sup/att* associations. By construction, each major claim is a root of a tree (called *ADU tree*), and all the trees created by the major claims together form the *ADU forest* of the article.

Given the ADU forest, the process for determining the stance of ADUs is visualised in Figure 6, and consists of the traversal of each ADU tree in isolation. Specifically, we initialise the stance of all major claims to have the stance value “for”. Then, each node (i.e., ADU) in the forest is traversed (root to leaves) to determine its stance as follows:

- If two nodes are connected with the *sup* relation, then the child node inherits the stance value of its parent.
- If two nodes are connected with the *att* relation, then we reverse the stance of the child node, compared to its parent, i.e., if the parent’s stance is “for”, the child’s stance becomes “against” and vice-versa.

D. Entity detection and linking

Our objective in this study is to help a journalist, or an interested citizen, better understand different facets of an issue or debate. Towards this, providing external, objective data and facts about important entities or concepts associated to the various arguments is important. As an example, in a discussion about an important public construction project, data and facts about the project itself, the contractor, or the public body responsible for the construction decisions might be relevant.

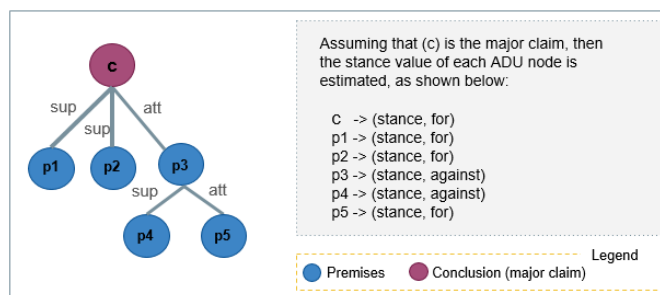


Figure 6. The stance-generation algorithm [2].

To achieve this, the ingestion process incorporates a *linking service* that enriches argumentative data with links to external sources from a variety of datasets from the Linked Open Data cloud (e.g., Wikidata, DBpedia, Wikipedia, etc).

In the ontology, each ADU node is connected to the *Entity Reference* class for identifying relevant named entities that refer to real-world events and associating them with external data sources (through the *Entity* class) using links to related articles, or other items that may help the reader assess their quality and trustworthiness [30]. The current implementation is based on two major steps: (i) extract substrings that are potential entity references, and (ii) identify the external resources (links) that describe the extracted entities. For more details on the following approach and results, see the work in Papantoniou [30].

The implementation process employs state-of-the-art tools, such as BERT [28], wikipedia2vec [31], and fastText [32]. More specifically, a publicly available BERT model is employed for training the *named entity recognizer* (step (i)) for the Greek language [29], while the vector representations of wikipedia2vec, fastText and BERT are used in the disambiguation process, a part of step (ii). The linked entity is selected over possibly many alternatives by calculating the similarity of candidate entity embeddings with the context of the ADU. This approach relies on the *distributional hypothesis* that words with similar meaning are usually found in a similar context.

The output of the linking process is a list of named entities (appearing in the ADUs found in the JSON), coupled with link(s) to related external data sources.

E. Cross-Document relationships detection

Argument mining also allows the identification of cross-document similarity relationships between ADUs appearing in different articles. In particular, each pair of ADUs is associated with a *similarity score*, recorded in the ontology using the class *Cross Document Relationship* and its associated properties. This score is leveraged to compute a similarity score among arguments, allowing users to identify similar arguments and find additional information related to a topic of interest.

V. USER-RELATED CONCEPTS IN ONTO4JARGS

In addition to the ingestion process, some parts of the DebateLab database are populated through user actions. These are described in the following subsections.

A. Quantitatively characterising arguments

The need for a quantitative characterization of the arguments' quality, acceptability, or other properties is important for a user who wants to better understand various facets of a debate. To achieve this, we employ the scoring algorithm *s-mDiCE* (*symmetric multi-Dimensional Comment Evaluation*) [33], with the aim to assess arguments along various dimensions (e.g., credibility, quality, acceptance, etc.).

The computation takes into account the relations among arguments, as well as votes that the users have placed on the arguments, represented using the class *Vote*. Each vote affects and is associated with a number of different aspects (such as informativeness, credibility, relevance, etc.) of the argument's evaluation process. A positive vote represents the fact that the user who submitted the vote agrees fully with the content of the argument. In the case of a negative vote, the user needs to specify the reasons he/she disagrees with the content of the argument, by choosing one or more aspects. This is similar to the approach used in other works (e.g., APOPSIS [34]).

Note that some of the evaluated aspects are static and are calculated once, whereas others are dynamically changing. The initial computation of scores takes place during the ingestion process, and recomputation is automatically performed when a relevant action takes place, i.e., a vote is placed by a user, or a new related argument is ingested.

B. Representing user profiles and activities

The *User* entity is used to represent user-related information for registered users interacting with the DebateLab ecosystem. Each user can edit his/her profile, which includes personal information, such as his/her display name (*USR_has_display_name*), year of birth (*UPR_year_of_birth*), registration date (*USR_registration_date*), and others.

In addition, the *User* entity records information related to the user's interaction with the system and personal preferences. More specifically, users can vote on arguments, directly affecting their evaluation (see Subsection V-A), and such votes are recorded using the *USR_submitted_argument_vote* property. In addition, each user can mark an article as a *favorite*, for easy access in the future (like a standard bookmarking service). Such articles are recorded using the *USR_has_favorite* property. A user can add and remove articles from his/her list of favorites, and can manage and organise this list, through the *Article Archiver* service of DebateLab. Furthermore, the user is allowed to request the ingestion of a new article in the DebateLab database, through the *on-demand article crawler* service. This has the effect of enriching the database with articles that the default crawling service has missed. Last but not least, the user can submit his/her own user-generated arguments through the Enhanced Debate Portal tool (similar to APOPSIS [34]), a debating platform for analysing structured opinions, integrated in the DebateLab ecosystem.

VI. CONCLUSION

The DebateLab project [1] aims to assist the professional journalist, as well as the interested citizen, to identify, under-

stand, analyse and navigate through arguments appearing in Greek journalistic articles, crawled from the Web. This can help support a better understanding of public debates, and to improved citizenship and e-democracy.

In this paper, we presented the knowledge model of DebateLab, i.e., our approach for representing and storing real arguments, extracted from existing sources (articles) on the Web. The representation approach is based on an ontology, called *Onto4JARGs*, that we built using Semantic Web Technologies based on the well-established AIF ontology [15]. We described the ontology implementation and its main components, the ingestion process (i.e., the process that transforms the textual raw data to structured knowledge represented in the ontology), and the allowed user actions that affect the contents of the DebateLab database.

Our ontology will be used in different scenarios for serving and helping the professional journalist in carrying out his/her daily activities more efficiently, through the implementation of DebateLab tools. Although our work is tailored for use in the DebateLab project (and the respective journalistic use case), we hope that it will be suitable for other scenarios and domains where argumentative information is relevant.

ACKNOWLEDGMENT

This work was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the "1st Call for H.F.R.I. Research Projects to support Faculty Members and Researchers and the procurement of high-cost research equipment" grant (Project Number: 4195).

REFERENCES

- [1] FORTH-ICS, "Debatelab project." [Online]. Available: <https://debatelab.ics.forth.gr/>, 2022, accessed: 2022-08-28.
- [2] ONTO4JARGs, "An rdf-based argumentation ontology for supporting journalistic arguments." [Online]. Available: https://isl.ics.forth.gr/debatelab_ontology/, 2022, accessed: 2022-08-28.
- [3] A. Singhal, "Introducing the knowledge graph: things, not strings," [Online]. Available: <https://blog.google/products/search/introducing-knowledge-graph-things-not/>, p. 16, 2012, accessed: 2022-08-28.
- [4] D. Allemang and H. James, *Semantic web for the working ontologist: effective modeling in RDFS and OWL*. San Francisco, CA, USA: Elsevier, 2011, ISBN 978-0-12-385965-5.
- [5] T. Berners-Lee, J. Hendler, and O. Lassila, "The semantic web," *Scientific American Magazine, USA*, vol. 284, no. 5, pp. 34–43, 2001.
- [6] S. Russel and P. Norvig, *Artificial intelligence: a modern approach*. Pearson Education Limited London, 2013, ISBN 978-0136042594.
- [7] T. Mikolov, K. Chen, G. S. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in *proceedings of the 1st International Workshop on Learning Representations (ICLR)*, Scottsdale, AZ, USA, 2013.
- [8] T. Mitchell *et al.*, "Never-ending learning," *Communications of the ACM*, vol. 61, no. 5, pp. 103–115, 2018.
- [9] N. L. Latar, "The robot journalist in the age of social physics: The end of human journalism?" in *The new world of transitioned media*. IDC, Herzliya, Israel: Springer, 2015, pp. 65–80, ISBN 978-3-319-09008-5.
- [10] A. Berven, O. A. Christensen, S. Moldeklev, A. L. Opdahl, and K. J. Villanger, "A knowledge-graph platform for newsrooms," *Computers in Industry*, vol. 123, p. 103321, 2020.
- [11] P. Castells, F. Perdrix, E. Pulido, M. Rico, R. Benjamins, J. Contreras, and J. Lorés, "Neptuno: Semantic web technologies for a digital newspaper archive," in *The Semantic Web: Research and Applications (ESWC)*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 445–458, ISBN 978-3-540-25956-5.

- [12] P. Vossen *et al.*, “Newsreader: Using knowledge resources in a cross-lingual reading machine to generate more knowledge from massive streams of news,” *Knowledge-Based Systems*, vol. 110, pp. 60–85, 2016.
- [13] M. Rospocher *et al.*, “Building event-centric knowledge graphs from news,” *Journal of Web Semantics*, vol. 37-38, pp. 132–151, 2016.
- [14] G. Leban, B. Fortuna, J. Brank, and M. Grobelnik, “Event registry: learning about world events from news,” in *Proceedings of the 23rd International Conference on World Wide Web*, Seoul Korea, 2014, pp. 107–110, ISBN 978-1-4503-2745-9.
- [15] C. Chesnevar *et al.*, “Towards an argument interchange format,” *The knowledge engineering review*, vol. 21, no. 4, pp. 293–316, 2006.
- [16] C. Chesnevar *et al.*, “The Argument Interchange Format (AIF) Specification,” *Knowledge Engineering Review (KER)*, p. 1–25, 01 2007.
- [17] University of Dundee, [Online]. AIF model. Available: <http://www.arg.dundee.ac.uk/aif/>, 2022, accessed: 2022-08-28.
- [18] P. Baroni, D. Gabbay, M. Giacomin, and L. van der Torre, *Handbook of Formal Argumentation*. College Publications, 2018, ISBN 978-1848902756.
- [19] Ph. Besnard and A. Hunter, “A review of argumentation based on deductive arguments,” in *Handbook of Formal Argumentation*, P. Baroni, D. Gabbay, M. Giacomin, and L. van der Torre, Eds. College Publications, 2018, pp. 435–482, ISBN 978-1848902756.
- [20] P. M. Dung, “On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games,” *Artificial Intelligence*, vol. 77, no. 2, pp. 321–357, September 1995.
- [21] RDF, “Resource description framework.” [Online]. Available: <https://www.w3.org/RDF/>, 2022, accessed: 2022-08-28.
- [22] A. Peldszus and M. Stede, “From argument diagrams to argumentation mining in texts: A survey,” *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI), USA*, vol. 7, no. 1, pp. 1–31, 2013.
- [23] D. Zografistou, G. Flouris, and D. Plexousakis, “Argql: A declarative language for querying argumentative dialogues,” in *RuleML+RR*. London, UK: Springer, 2017, pp. 230–237, ISBN 978-3-319-61251-5.
- [24] D. Zografistou, G. Flouris, T. Patkos, and D. Plexousakis, “Implementing the argql query,” in *Computational Models of Argument - Proceedings of COMMA*, vol. 305. Warsaw, Poland: IOS Press, 2018, pp. 241–248.
- [25] J. Lawrence and C. Reed, “Argument mining: A survey,” *Computational Linguistics, Cambridge, USA*, vol. 45, no. 4, pp. 765–818, 2020, ISBN 0891-2017.
- [26] R. M. Palau and M. F. Moens, “Argumentation mining: the detection, classification and structure of arguments in text,” in *Proc. of the 12th ICAIL*, Barcelona, Spain, 2009, pp. 98–107, ISBN 978-1-60558-597-0.
- [27] A. Vaswani *et al.*, “Attention is all you need,” in *Advances in neural information processing systems (NIPS)*, vol. 30, CA, USA, 2017, p. 6000–6010.
- [28] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: pre-training of deep bidirectional transformers for language understanding,” *Proceedings of the 2019 Conference of the Association for Computational Linguistics: Human Language Technologies (NAACL), Minneapolis, Minnesota, USA*, vol. 1, pp. 4171–4186, 2019.
- [29] J. Koutsikakis, I. Chalkidis, P. Malakasiotis, and I. Androutsopoulos, “Greek-bert: The greeks visiting sesame street,” in *11th Hellenic Conference on Artificial Intelligence (SETN)*. New York, NY, USA: Association for Computing Machinery, 2020, pp. 110–117, ISBN 978-1-4503-8878-8.
- [30] K. Papantoniou, V. Efthymiou, and G. Flouris, “El-nel: Entity linking for greek news articles,” in *Proceedings of the ISWC*, vol. 2980. Virtual Conference: CEUR-WS.org, 2021.
- [31] I. Yamada *et al.*, “Wikipedia2vec: An efficient toolkit for learning and visualizing the embeddings of words and entities from wikipedia,” in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (EMNLP)*. [Online]: Association for Computational Linguistics, 2020, p. 23–30.
- [32] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, “Enriching word vectors with subword information,” *Transactions of the Association for Computational Linguistics (TACL), Vancouver, Canada*, vol. 5, pp. 135–146, 2017.
- [33] T. Patkos, G. Flouris, and A. Bikakis, “Symmetric multi-aspect evaluation of comments extended abstract,” in *In Proceedings of the 22nd European Conference on Artificial Intelligence (ECAI)*. Amsterdam, Netherlands: IOS Press, 2016, pp. 1672–1673, ISBN 978-1-61499-671-2.
- [34] E. Ymeralli, G. Flouris, T. Patkos, and D. Plexousakis, “Apopis: A web-based platform for the analysis of structured dialogues,” in *The 16th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE)*. Rodos, Greece: Springer, 2017, pp. 224–241, ISBN 978-3-319-69458-0.

Towards Exchanging Wearable-PGHD with EHRs: Developing a Standardized Information Model for Wearable-Based Patient Generated Health Data

Abdullahi Abubakar Kawu
School of Computer Science,
Technological University, Dublin
Dublin, Ireland
Email: d21125967@mytudublin.ie

Dympna O’Sullivan
School of Computer Science,
Technological University, Dublin
Dublin, Ireland
Email:
dympna.osullivan@tudublin.ie

Lucy Hederman
School of Computer Science and
Statistics,
Trinity College Dublin
Dublin, Ireland
Email: hederman@tcd.ie

Abstract— Wearables have become commonplace for tracking and making sense of patient lifestyle, wellbeing and health data. Most of this tracking is done by individuals outside of clinical settings, however some data from wearables may be useful in a clinical context. As such, wearables may be considered a prominent source of Patient Generated Health Data (PGHD). Studies have attempted to maximize the use of the data from wearables including integrating with Electronic Health Records (EHRs). However, usually a limited number of wearables are considered for integration and, in many cases, only one brand is investigated. In addition, we find limited studies on integration of metadata including data quality and provenance, despite such data being very relevant for clinical decision making. This paper describes a proposed design and development of a generic information model for wearable based PGHD integration with EHRs. We propose a vendor-neutral model that can work with a wider range of wearables and discuss our proposed method to employ an ontology-based approach and provide insights to future work.

Keywords-wearables; electronic health records; interoperability; information model; FHIR; fast healthcare interoperability resources.

I. INTRODUCTION

Patient Generated Health Data (PGHD) refers to health data that a patient (or their authorized representative) records outside the clinic setting, is relevant to their wellbeing and can be used by them or clinicians for their health management. PGHD is collected using many mediums including Patient Health Records (PHRs), mobile health application (mHealth), and wearables. Wearables are a prominent source of PGHD identified in literature. Wearables can collect various granular types of data using inbuilt sensors. These devices combine sensing capabilities with algorithms to produce data in both raw and aggregated formats which can be effective for healthcare monitoring (by both the patient and clinicians) [1]. Data from sensors is used to measure various data – for example, heart rate, daily number of steps taken, distance covered, number of calories burned, number of floors climbed. From these sensors, data such as detailed heart rate history with heart rate zones, active minutes and sleep duration and quality can also be calculated and can be viewed through the associated web-

based application or mobile app, and for a limited period, on the device too. For most wearables, the internal memory of the device is only able to store minute-by-minute data of the last seven days, and 30 days of daily summaries [2]. Other data can also be collected from the trackers too, such as last sync date, battery level, etc. [3]. Wearables provide an opportunity for tracking patients’ health condition in their regular living settings, providing insightful data about a patient, more complete than what can be collected during infrequent clinical visits. Globally, the number of wearable devices has increased from 325 million in 2016 to 722 million in 2019, more than doubling in only three years [4]. By 2022, there will likely be more than one billion of such devices worldwide. Vendrico [5] have curated an information database of 431 wearable devices produced by 266 companies; there are potentially other bespoke ones that are unaccounted for in this estimate. The use of wearables for patient care has benefits such as: connected information, patient-oriented healthcare, and gamification [6]. For the purposes of this study, our definition of wearable did not include belt-based wearables as this has been found to be unappealing to users [7].

Due to benefits and prospects of PGHD for personalized care and population health, there is significant interest and investment in integrating PGHD with electronic health records. Jung [8] integrated PHRs with EHR using lifelogs, but suggested that wearables could help in reporting more objective data, with less burden to the patient. Similarly, Plastiras and O’Sullivan [9] developed an information-model for integrating PGHD and Observation of Daily Living (ODL) with EHR. However, they considered only one wearable in their study – a Fitbit tracker. We argue that this may not be generalizable to other wearables, as there could be more data and meta data that is unaccounted for in this approach. Also, there is less consideration for meta data on data quality such as device accuracy, which is of essence in the decision-making process of a clinician. In their study of four commercial widely used wearables, Kaewkannate and Kim [10], established that Fitbit Flex and Misfit have difficulty in detecting when a user climbs or descends stairs. Despite Fitbit leading in the consumer market of wrist-worn

wearables [11], the Fitbit Flex model should not be relied upon for use in patient climbing or descending tasks. Reporting similar concern for other wearable-based data, Wood, Bennett, and Basch [12] state that it is not clear whether sleep data or any other data from one wearable can be interchangeable with another and whether this will result to the same meaning of the data. The aforementioned points make metadata information very important for clinical decision making, hence our interest in providing valuable metadata alongside sensor data to clinicians.

This short position paper proposes a generic framework to enable a wide range of wearable-PGHD to be interoperable with EHRs, to allow seamless exchange of clinically relevant data from patients to providers. The framework will enable a class of wearables to be integrated with EHR systems using an ontology-driven Information Model (IM) based on Fast Health Interoperability Resources (FHIR). In this paper, we provide a description of our proposed IM for transferring information in a standardized way between wearables and EHR systems. In the rest of this paper, we describe our proposed method to develop the interfacing layer between EHR and wearables as well as present insight into our proposed future work. In Section II, we describe the proposed methodology and architecture, outlining stages and steps to be carried out to design the proposed model, including the underlining technologies to be employed. Section III presents future work on the proposed design, and Section IV gives a summary and a conclusion on the position paper.

II. PROPOSED METHODOLOGY AND ARCHITECTURE

Plastiras, and O'Sullivan [9] and Plastiras, O'Sullivan, and Weller [13] describe steps for designing and developing an ontology-driven information model for EHR integration. This includes, but may not be limited to, analysis of common functionality and data to determine information to be exchanged, review of standards, and developing a middleware for document exchange. Similar to [9], our research proposes the use of an ontology-driven IM to address issues of semantic and syntactic interoperability between wearable-based PGHD and EHR systems. Figure 1 below depicts our proposed architecture for wearable-based PGHD-EHR integration based on FHIR standard. The following section describes the proposed architecture and future steps and thoughts around our proposal for developing ontology-driven IM, to be derived directly from common wearable-based PGHD and functions.

As previously mentioned, most wearables have a separate persistent long-term data staging, visualization and storage platform, due to their limited memory and battery size. So, in addition to owning a wearable device, most users have an accompanying mobile application that they use alongside the wearable, to enable them to visualize and store data. In addition, they can sync and have access to their data online. With this structure, wearable data can be accessed, manipulated and shared only through the web or mobile application. In our architecture, we have embodied this, but in addition, we depict data sharing capabilities that can be explored for integration (Figure 1, Stage 1 - 3). Wearable-based PGHD can be shared with an EHR system using standard clinical document exchange format such as Clinical Care Document (CCD), Clinical Care Record (CCR), and XML [14]. However, because FHIR is the most recent standard from Health Level 7 (HL7)[15] that overcomes the shortcomings of the previously mentioned document exchange standards, we intend to employ FHIR as the messaging standard in our proposed design. FHIR has been previously employed to exchange some wearable based data with an EHR application [14], [16]. However, most of the studies have relied on the use of data from one specific brand or device, which may not make it generalizable.

In our design, we propose that whenever wearable-based PGHD is to be shared with clinicians (providers), it will go via a middleware interface that will transform the data into FHIR-ready data, a procedure we refer to as FHIRification (Figure 1, Stage 4). At this layer, data will also be mapped to standard and custom ontologies present at that interface, which can fit to the terminology component of FHIR. The FHIRification involves transforming data to fit into existing FHIR Resources and Extensions. The transformation engine should also generate a FHIR relevant Conformance and Capability Statement that documents the required server implementation of the FHIR resources. Thereafter, data is made available to the EHR for use (Figure 1, Stage 5). Similar to the approach in many studies [14], we propose that demographic data related to patients are transformed into the Patient resource, data related to actual sensor data is fitted into the Encounter resource, and data related to the device are transformed into the Device resource of FHIR. Derived data from patient's medical history can also be considered under Condition resource. In addition, we intend to explore the use of Provenance resource in integration [17], to help document data provenance. We also posit that a PGHD profile will be desirable to provide a container for all PGHD resources, for ease of reference, use and management.

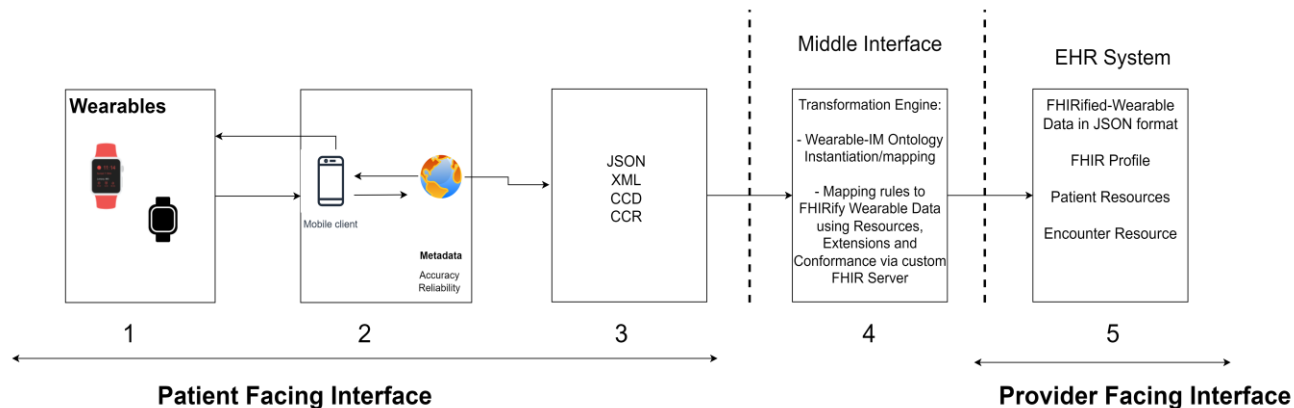


Figure 1. Proposed Wearable-based PGHD-EHR integration

With emerging diseases and in preparation for another pandemic, more wearables with new capabilities are churned out, leading to newer forms of datasets and features and data that are bespoke. Data such as temperature sensing (relating to women’s health), ECG (for atrial fibrillation monitoring) and skin temperature etc, are evolving and need to be considered. Hence, our approach is to make this consideration of this diversity in data from these wearables to enable wider adoption and use of the framework.

III. FUTURE WORK

Information Model development is carried out in stages which include identification, and evaluation of available data and functionality, determining candidate data to be exchanged, leveraging existing EHR standards required for candidate data syntactic integration, and the design of an ontology for semantic integration [9]. Hence, more work is expected to fully develop the system and demonstrate the effectiveness. However, here we outline steps to be undertaken towards our proposed ontology-driven IM, and how we will leverage other state of art like the use of FHIR standards.

A. Analysis of common wearable functionality and features

We propose to review the features and functionality of wearables based on the work of Kaekwannate and Kim [10] who identify common data and functions of four popular wearable devices. In their studies, they outline features of JawboneUp24, Fitbit Flex, Withings Pulse, Misfit Shine wearables [10]. However, in addition to this, more wearables will be identified from [5] to be included for comparison and evaluation. Wearables like Apple Watch and Xaiomi are included in the list of wearables to be compared for the purpose of common data and functions identification. While generalizable categories of data such as Activity Tracking Data (Distance, Calories, Floors climbed, Intensity), Sleep Data (Duration, Stage, Score), Stress Data and other known health data from wearables

will be prioritized, we will also consider bespoke data. A scoring system will be employed to assign a utility score to each wearable based on a select criterion, to help us identify generalizable data, but also bespoke data and data with the most value.

B. EHR Data Exchange standards and interoperability

Recent work by Microsoft [18] focused on achieving interoperability by developing a platform to share historical patient data from Fitbit into FHIR server. We will learn from this work but extend it to embrace other data and functions that may come from other wearables, and data that could enhance Fitbit’s overall value within a PGHD-EHR ecosystem, such as accuracy and reliability metadata. In our proposed architecture, we intend to pay attention to data quality using metadata and provenance, to improve clinicians’ confidence in PGHD. Data accuracy, reliability validity, and completeness are foremost data quality issues that have hindered PGHD integration [19], [20]. In our proposed system, data exchanged between wearables and EHRs must conform to relevant structure and syntactic rules. The syntactic rules will closely align with FHIR standards, and in that case, information can be transformed to and transferred as a JSON document. However, it will also be transformable to legacy standards such as XML.

C. Design of Wearable-based PGHD Ontology

We intend to employ open source software Protégé for developing a wearable-based PGHD ontology, and OWL to instantiate the ontology at the middle interface [13]. Using an ontology-based approach, semantic interoperability issues can be avoided. The usage of different terms by wearables and EHRs to define the same concept can prevent data sharing with EHRs. Data such “Oxygen Saturation Measurement” in one wearable can be represented as “SpO2” in another wearable, hence a standardize ontology that will fit appropriately to the Terminology component of FHIR becomes desirable. The ontology will define relationship, constraints, and concepts about the data from the wearables. Similarly, standard coding schemas will be employed too. For instance, the SNOMED CT code for

SpO2 is “431314004”. By assigning this code to Oxygen Saturation Measurement, the meaning of these two data could be interpreted as same.

IV. CONCLUSION

Wearables are digital health products or devices that are being used to collect, store and process health data, also known as PGHD, towards providing holistic digital health information that can be used for patient monitoring or to encourage or enact behavior change. Data collected by wearables are used across a broad set of health domains and towards population health, becoming a key contributor to digital health initiatives. They are often combined with mobile and web applications to process and manage data. In this paper, we present preliminary work in designing an ontology-based IM for wearable-based PGHD for integration with EHR. The proposed IM is proposed to be derived from common features and functionality of a wide range of consumer-grade wearables and will employ the FHIR standard for interoperability with EHR. We outline steps required to develop this fully in the future and discussed how this can be implemented.

ACKNOWLEDGMENT

This work was conducted with the financial support of the Science Foundation Ireland Centre for Research Training in Digitally Enhanced Reality (d-real) under Grant No. 18/CRT/6224 and at the ADAPT SFI Research Centre for AI-Driven Digital Content Technology under Grant No. 13/RC/2106_P2 at Technological University Dublin. For the purpose of Open Access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

REFERENCES

- [1] A. Godfrey, V. Hetherington, H. Shum, P. Bonato, N. Lovell, and S. Stuart, “From A to Z: Wearable technology explained,” *Maturitas*, vol. 113, pp. 40–47, 2018.
- [2] M. Schellevis, B. Jacobs, C. Meijer, and J. de Ruiter, “Getting access to your own Fitbit data,” *Radboud University*, 2016.
- [3] G. Crenshaw, “FitBit API,” *The ABCs of Fitbit APIs*, Jul. 11, 2022. https://pages.fitbit.com/2021_11-WebinarVideo-TheABCsofFitbitAPIs_21-VideoReplay.html?aliId=eyJpIjoiUlhncUIZUHU0SWVpMWR6MiIsInQiOiJ2TVd0VEhGS05cL0JBRzc4Zk80d0oyQT09In0%253D (accessed Jul. 11, 2022).
- [4] F. Larichhia, “Global Connected Devices Worldwide 2022,” *Global Connected Devices Worldwide 2022*, 02 2022. <https://www.statista.com/statistics/487291/global-connected-wearable-devices/> (accessed Jun. 07, 2022).
- [5] Vendrico, “The Wearables Database,” *Wearable Device Database*, Jul. 08, 2022. <https://vandrico.com/wearables.html> (accessed Jul. 08, 2022).
- [6] S. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K.-S. Kwak, “The internet of things for health care: a comprehensive survey,” *IEEE access*, vol. 3, pp. 678–708, 015.
- [7] I. Bendich, “Changes in prospectively collected longitudinal patient-generated health data are associated with short-term patient-reported outcomes after total joint arthroplasty: a pilot study,” *Arthroplasty Today*, vol. 5, no. 1, pp. 61–63, 2019, doi: 10.1016/j.artd.2019.01.005.
- [8] S. Y. Jung, “Development of comprehensive personal health records integrating patient-generated health data directly from samsung s-health and apple health apps: Retrospective cross-sectional observational study,” *JMIR mHealth and uHealth*, vol. 7, no. 5, 2019, doi: 10.2196/12691.
- [9] P. Plastiras and D. O’Sullivan, “Exchanging personal health data with electronic health records: A standardized information model for patient generated health data and observations of daily living,” *International journal of medical informatics*, vol. 120, pp. 116–125, 2018.
- [10] K. Kaewkannate and S. Kim, “A comparison of wearable fitness devices,” *BMC public health*, vol. 16, no. 1, pp. 1–16, 2016.
- [11] A. Ng, R. Kornfield, S. M. Schueller, A. K. Zalta, M. Brennan, and M. Reddy, “Provider perspectives on integrating sensor-captured patient-generated data in mental health care,” *Proceedings of the ACM on human-computer interaction*, vol. 3, no. CSCW, pp. 1–25, 2019.
- [12] W. A. Wood, A. V. Bennett, and E. Basch, “Emerging uses of patient generated health data in clinical research,” *Molecular oncology*, vol. 9, no. 5, pp. 1018–1024, 2015.
- [13] P. Plastiras, D. O’Sullivan, and P. Weller, “An ontology-driven information model for interoperability of personal and electronic health records,” 2014.
- [14] R. K. Saripalle, “Fast Health Interoperability Resources (FHIR): current status in the healthcare system,” *International Journal of E-Health and Medical Communications (IJEHMC)*, vol. 10, no. 1, pp. 76–93, 2019.
- [15] D. Bender and K. Sartipi, “HL7 FHIR: An Agile and RESTful approach to healthcare information exchange,” presented at the Proceedings of the 26th IEEE international symposium on computer-based medical systems, 2013, pp. 326–331.
- [16] J. Ruminski, A. Bujnowski, T. Kocejko, A. Andrushevich, M. Biallas, and R. Kistler, “The data exchange between smart glasses and healthcare information systems using the HL7 FHIR standard,” presented at the 2016 9th International Conference on Human System Interactions (HSI), 2016, pp. 525–531.
- [17] J. Moehrke, A. Rosenthal, and A. Chapman, “Extending the FHIR standard to handle provenance,” *The MITRE Corporation*, 2016.
- [18] Microsoft, “FitBit on FHIR.” Accessed: Jul. 06, 2022. [Online]. Available: <https://github.com/microsoft/FitbitOnFHIR>
- [19] J. Codella, C. Partovian, H.-Y. Chang, and C.-H. Chen, “Data quality challenges for person-generated health and wellness data,” *IBM Journal of Research and Development*, vol. 62, no. 1, pp. 3–1, 2018.
- [20] R. Abdolkhani, “Patient-generated health data management and quality challenges in remote patient monitoring,” *JAMIA Open*, vol. 2, no. 4, pp. 471–478, 2019, doi: 10.1093/jamiaopen/ooz036.