|  |
| --- |
| Nr 1 |
| Tipologia prodotto Contributo in Conferenza o Convegno (Senza ISBN) |
| Titolo OpeNER and PANACEA: Web Services for the CLARIN Research Infrastructure |
| Descrizione |
| Elenco autori Davide Albanesi Riccardo Del Gratta |
| Ruolo svolto |
| anno pubblicazione 2019 |
| abstract: This paper describes the necessary steps for the integration of OpeNer and PANACEA Web Services within the CLARIN research infrastructure. The original Web Services are wrapped into a framework and re-implemented as REST APIs to be further exploited through both Language Resource Switchboard and WebLicht and made available for the CLARIN community.In url: https://office.clarin.eu/v/CE-2019-1512\_CLARIN2019\_ConferenceProceedings.pdf abstract: This paper describes the necessary steps for the integration of OpeNer and PANACEA Web Services within the CLARIN research infrastructure. The original Web Services are wrapped into a framework and re-implemented as REST APIs to be further exploited through both Language Resource Switchboard and WebLicht and made available for the CLARIN community. |

|  |
| --- |
| Nr 2 |
| Tipologia prodotto Articolo in Rivista |
| Elenco autori Sprugnoli R.; Pardelli G.; Boschetti F.; Del Gratta R. |
| Titolo Un'Analisi Multidimensionale della Ricerca Italiana nel Campo delle Digital Humanities e della Linguistica Computazionale |
| Rivista Umanistica Digitale |
| Codice identificativo (ISSN) 2532-8816 |
| anno pubblicazione 2019 |
| Indice di classificazione |
| Impact Factor rivista |
| ruolo svolto |
| numero citazioni |
| Altre informazioni doi: 10.6092/issn.2532-8816/8581 url: https://umanisticadigitale.unibo.it/article/view/8581 abstract: This article proposes the first comparative study of four years of Italian conferences in the fields of Digital Humanities and Computational Linguistics. More specifically, we created a corpus consisting of the contributions presented in the AIUCD and CLiC-it conferences between 2014 and 2017 to which we applied a multidimensional analysis taking into consideration: (i) the study of collaborations between authors using social networks analysis techniques, (ii) the automatic extraction of terminology and information and (iii) the examination of citational practices. By combining both qualitative and quantitative methods of investigation, this paper aims to shed light on convergences and discrepancies between two research areas that historically have common origins. |

|  |
| --- |
| Nr 3 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Lionel Nicolas, Alexander König, Monica Monachini, Riccardo Del Gratta, Silvia Calamai, Andrea Abel, Alessandro Enea, Francesca Biliotti, Valeria Quochi, Francesco Vincenzo Stella |
| Titolo CLARIN-IT: State of Affairs, Challenges and Opportunities |
| Codice identificativo (ISBN) 978-91-7685-273-6 |
| anno pubblicazione 2018 |
| abstract: his paper gives an overview on the Italian national CLARIN consortium as it currently stands two years after its creation at the end of 2015. It thus discusses the current state of affairs of the consortium on several aspects, especially with regards to members. It also discusses the events and initiatives that have been undertaken, as well as the ones that are planned in the close future. It finally outlines the conclusions of a user survey performed to understand the expectations of a targeted user population and provides indications regarding the next steps planned.Selected papers from the CLARIN Annual Conference 2017, Budapest, 18-20 September 2017 url: http://www.ep.liu.se/ecp/contents.asp?issue=147 abstract: his paper gives an overview on the Italian national CLARIN consortium as it currently stands two years after its creation at the end of 2015. It thus discusses the current state of affairs of the consortium on several aspects, especially with regards to members. It also discusses the events and initiatives that have been undertaken, as well as the ones that are planned in the close future. It finally outlines the conclusions of a user survey performed to understand the expectations of a targeted user population and provides indications regarding the next steps planned. |

|  |
| --- |
| Nr 4 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Riccardo Del Gratta, Sara Goggi, Gabriella Pardelli, Nicoletta Calzolari |
| Titolo LREMap, a Song of Resources and Evaluation |
| Codice identificativo (ISBN) 979-10-95546-00-9 |
| anno pubblicazione 2018 |
| abstract: After 8 years we revisit the LRE Map of Language Resources, introduced at LREC 2010, to try to get a picture of the field and its evolution as reflected by the creation and use of Language Resources. The purpose of the Map was in fact "to shed light on the vast amount of resources that represent the background of the research presented at LREC". It also aimed at a "change of culture in the field, actively engaging each researcher in the documentation task about resources". The data analysed here have been provided by the authors of several conferences during the phase of submission of papers, and contain information about ca. 7500 resources. We analysed the LRE Map data from many different viewpoints and the paper reports on the global picture, on different trends emerging from the diachronic perspective and finally on some comparisons between the 2 major conferences present in the Map: LREC and COLING.Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018) url: http://www.lrec-conf.org/proceedings/lrec2018/summaries/300.html abstract: After 8 years we revisit the LRE Map of Language Resources, introduced at LREC 2010, to try to get a picture of the field and its evolution as reflected by the creation and use of Language Resources. The purpose of the Map was in fact \"to shed light on the vast amount of resources that represent the background of the research presented at LREC\". It also aimed at a \"change of culture in the field, actively engaging each researcher in the documentation task about resources\". The data analysed here have been provided by the authors of several conferences during the phase of submission of papers, and contain information about ca. 7500 resources. We analysed the LRE Map data from many different viewpoints and the paper reports on the global picture, on different trends emerging from the diachronic perspective and finally on some comparisons between the 2 major conferences present in the Map: LREC and COLING. |

|  |
| --- |
| Nr 5 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Luigi Bambaci, Federico Boschetti, Riccardo Del Gratta |
| Titolo Qohelet Euporia: a Domain Specific Language to Annotate Multilingual Variant Readings |
| Codice identificativo (ISBN) 978-1-5386-4385-3 |
| anno pubblicazione 2018 |
| abstract: Encoding multilingual variant readings is time-consuming and error-prone. The guidelines provided by the Text Encoding Initiative (TEI) ensure data interchange, but the TEIXML verbosity risks to distract annotators with a traditional background in philological studies from their critical activity. We illustrate how a Domain Specific Language (DSL) facilitates both the manual annotation of the critical apparatus and the data interchange. Our case study is based on the multilingual tradition of the biblical book of Qohelet, which has been annotated through the annotation tool based on DSLs named Euporia. abstract: Encoding multilingual variant readings is time-consuming and error-prone. The guidelines provided by the Text Encoding Initiative (TEI) ensure data interchange, but the TEIXML verbosity risks to distract annotators with a traditional background in philological studies from their critical activity. We illustrate how a Domain Specific Language (DSL) facilitates both the manual annotation of the critical apparatus and the data interchange. Our case study is based on the multilingual tradition of the biblical book of Qohelet, which has been annotated through the annotation tool based on DSLs named Euporia. |

|  |
| --- |
| Nr 6 |
| Tipologia prodotto Contributo in Conferenza o Convegno (Senza ISBN) |
| Titolo (Re)Using OpeNER and PANACEA Web Services in the CLARIN Research Infrastructure |
| Descrizione |
| Elenco autori Del Gratta Riccardo |
| Ruolo svolto |
| anno pubblicazione 2017 |
| abstract: We describe the implications of (re)using the OpeNer and PANACEA Web Services into the CLARIN Research Infrastructure. The analyzed tools are of great interest for specific communities such as academic and small business focused on sentiment/opinion analysis and on Machine Translation along with related technologies, but their outcomes may be of great importance for the CLARIN audience as well. In fact, the Virtual Language Observatory shows a lot of lexical resources for sentiment but a few tool, while a lot of lexical resources and tools are available for Machine Translation. This means that the latter community is already in CLARIN, while the former should be poked. If community-related challenges are on the political side, issues related to interoperability are definitely on the technical one. The initiative is carried out at the ILC4CLARIN center in Pisa, the leading one of the CLARIN-IT national Consortium. The least common multiple between those two projects is neither limited to tools and Web Services nor to the creation of annotated corpora and lexicons; neither to the focus they have on specific communities. They also are based on (and strongly pursue and suggest) the concept of interoperability. This is clear from the use of the Kyoto Annotation Format in OpeNer, of Graph Annotation Format in PANACEA8 and of and the Lexical Markup Framework in both. Data and tools interoperability is also a key asset in both CLARIN (https://www.clarin.eu/event/2017/clarin-workshop-towards-interoperability-lexico-semantic-resources) and EUDAT (https://eudat.eu/communities/an-eudat-based-fair-data-approach-for-data-interoperability) . Within CLARIN, initiatives such as the Language Resource Switchboard and openly go towards methodologies and "systems" to address the interoperability issues. From a technical point of view the main issues are briefly reported below: 1. Many tools in OpeNer and PANACEA are command line ones; 2. OpeNer o\_ers both POST and GET API; 3. PANACEA built its Web Services using Soaplab11 and o\_ers SOAP Web Services; 4. KAF, LMF and GrAF guarantee the interoperability among data and services; 5. Simple pipelines are available in OpeNer, while a workow engine has been used in PANACEA. Tools are already wrapped, but to fully meet the requirements of both LRS and WebLicht we have to build a new shell around the command line tools so that REST APIs can accept both POST and GET requests and accept/produce different formats. Indeed if Language Resource Switchboard accepts tools with their output format but requires to read data from URL in plain text, WebLicht accepts tools which read and write the TCF format. While OpeNer requires that the core (the command line) be wrapped into a REST shell, Web Services in PANACEA need REST APIs around a SOAP core. In the final paper, we will finalize the technical aspects and describe how the User Involvement group can play an important role in poking the sentiment/opinion community in CLARIN.In url: https://indico.egi.eu/indico/event/3455/contribution/139 abstract: We describe the implications of (re)using the OpeNer and PANACEA Web Services into the CLARIN Research Infrastructure. The analyzed tools are of great interest for specific communities such as academic and small business focused on sentiment/opinion analysis and on Machine Translation along with related technologies, but their outcomes may be of great importance for the CLARIN audience as well. In fact, the Virtual Language Observatory shows a lot of lexical resources for sentiment but a few tool, while a lot of lexical resources and tools are available for Machine Translation. This means that the latter community is already in CLARIN, while the former should be poked. If community-related challenges are on the political side, issues related to interoperability are definitely on the technical one. The initiative is carried out at the ILC4CLARIN center in Pisa, the leading one of the CLARIN-IT national Consortium. The least common multiple between those two projects is neither limited to tools and Web Services nor to the creation of annotated corpora and lexicons; neither to the focus they have on specific communities. They also are based on (and strongly pursue and suggest) the concept of interoperability. This is clear from the use of the Kyoto Annotation Format in OpeNer, of Graph Annotation Format in PANACEA8 and of and the Lexical Markup Framework in both. Data and tools interoperability is also a key asset in both CLARIN (https://www.clarin.eu/event/2017/clarin-workshop-towards-interoperability-lexico-semantic-resources) and EUDAT (https://eudat.eu/communities/an-eudat-based-fair-data-approach-for-data-interoperability) . Within CLARIN, initiatives such as the Language Resource Switchboard and openly go towards methodologies and \"systems\" to address the interoperability issues. From a technical point of view the main issues are briefly reported below: 1. Many tools in OpeNer and PANACEA are command line ones; 2. OpeNer o\_ers both POST and GET API; 3. PANACEA built its Web Services using Soaplab11 and o\_ers SOAP Web Services; 4. KAF, LMF and GrAF guarantee the interoperability among data and services; 5. Simple pipelines are available in OpeNer, while a workow engine has been used in PANACEA. Tools are already wrapped, but to fully meet the requirements of both LRS and WebLicht we have to build a new shell around the command line tools so that REST APIs can accept both POST and GET requests and accept/produce different formats. Indeed if Language Resource Switchboard accepts tools with their output format but requires to read data from URL in plain text, WebLicht accepts tools which read and write the TCF format. While OpeNer requires that the core (the command line) be wrapped into a REST shell, Web Services in PANACEA need REST APIs around a SOAP core. In the final paper, we will finalize the technical aspects and describe how the User Involvement group can play an important role in poking the sentiment/opinion community in CLARIN. |

|  |
| --- |
| Nr 7 |
| Tipologia prodotto Contributo in Conferenza o Convegno (Senza ISBN) |
| Titolo AIUCD e CLiC-it : citazioni bibliografiche a confronto |
| Descrizione |
| Elenco autori Gabriella Pardelli, Silvia Giannini, Federico Boschetti, Riccardo Del Gratta |
| Ruolo svolto |
| anno pubblicazione 2017 |
| abstract: Il lavoro propone l'analisi e il confronto dei riferimenti bibliografici delle cinque edizioni annuali della Conferenza dell'Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD) e del primo biennio (2014-2015) della Conferenza Italiana di Linguistica Computazionale (CLiC-it) per misurare la direzione in cui si muove il trend citazionale.. L'analisi muove dal principio di rilevanza della citazione nella trasmissione della conoscenza in un periodo di grandi cambiamenti socioculturali e di importanti evoluzioni nelle modalità di produzione e diffusione dei risultati della ricerca scientifica. Lo scopo dello studio è quello di misurare la gamma delle risorse citate in questa area del sapere mediante l'uso di indicatori volti a comprendere la loro ampiezza, l'estensione temporale, la varietà, le relazioni con il mondo editoriale e i modelli di riferimento. L'osservazione dei risultati consente di classificare i documenti citati, di descriverne le caratteristiche e di valutare eventuali trasformazioni rispetto alle modalità di citazione tradizionali. The paper suggests the analysis of the bibliographic references - enclosed in the proceedings of the Italian Conference on Computational Linguistics - CLICit in 2014-2015 biennium and of five AIUCD Conference editions 2012-2016 . The analysis moves from the principle of relevance of citation in the transmission of knowledge in a period of great socio-cultural changes and important developments in the production and dissemination of the results in the scientific research. The purpose of the study is to measure the range of resources cited in this area of knowledge by the use of indicators aimed in understanding their wideness, the time extension, the variety, the relations with the publishing world and the reference models. The observation of the results allows to classify the cited document, to describe its characteristics and to assess any changes compared to the traditional citation mode.In AIUCD 2017 Conference url: http://aiucd2017.aiucd.it/wp-content/uploads/2017/01/book-of-abstract-AIUCD-2017.pdf abstract: Il lavoro propone l\'analisi e il confronto dei riferimenti bibliografici delle cinque edizioni annuali della Conferenza dell\'Associazione per l\'Informatica Umanistica e la Cultura Digitale (AIUCD) e del primo biennio (2014-2015) della Conferenza Italiana di Linguistica Computazionale (CLiC-it) per misurare la direzione in cui si muove il trend citazionale.. L\'analisi muove dal principio di rilevanza della citazione nella trasmissione della conoscenza in un periodo di grandi cambiamenti socioculturali e di importanti evoluzioni nelle modalità di produzione e diffusione dei risultati della ricerca scientifica. Lo scopo dello studio è quello di misurare la gamma delle risorse citate in questa area del sapere mediante l\'uso di indicatori volti a comprendere la loro ampiezza, l\'estensione temporale, la varietà, le relazioni con il mondo editoriale e i modelli di riferimento. L\'osservazione dei risultati consente di classificare i documenti citati, di descriverne le caratteristiche e di valutare eventuali trasformazioni rispetto alle modalità di citazione tradizionali. The paper suggests the analysis of the bibliographic references - enclosed in the proceedings of the Italian Conference on Computational Linguistics - CLICit in 2014-2015 biennium and of five AIUCD Conference editions 2012-2016 . The analysis moves from the principle of relevance of citation in the transmission of knowledge in a period of great socio-cultural changes and important developments in the production and dissemination of the results in the scientific research. The purpose of the study is to measure the range of resources cited in this area of knowledge by the use of indicators aimed in understanding their wideness, the time extension, the variety, the relations with the publishing world and the reference models. The observation of the results allows to classify the cited document, to describe its characteristics and to assess any changes compared to the traditional citation mode. |

|  |
| --- |
| Nr 8 |
| Tipologia prodotto Contributo in Conferenza o Convegno (Senza ISBN) |
| Titolo CLARIN-IT: State of Affairs, Challenges and Opportunities |
| Descrizione |
| Elenco autori Lionel Nicolas, Alexander Konig, Monica Monachini, Riccardo Del Gratta, Silvia Calamai, Andrea Abel, Alessandro Enea, Francesca Biliotti, Valeria Quochi |
| Ruolo svolto |
| anno pubblicazione 2017 |
| abstract: This paper provides an overview on the Italian national CLARIN consortium and the status of CLARIN-IT in general. It thus discusses the current state of affairs of the consortium and provi-des information on the members, especially with regards to what they offer to CLARIN in terms of resources, services and expertise, and what CLARIN offers them to further their own research.In url: https://www.clarin.eu/event/2017/clarin-annual-conference-2017-budapest-hungary abstract: This paper provides an overview on the Italian national CLARIN consortium and the status of CLARIN-IT in general. It thus discusses the current state of affairs of the consortium and provi-des information on the members, especially with regards to what they offer to CLARIN in terms of resources, services and expertise, and what CLARIN offers them to further their own research. |

|  |
| --- |
| Nr 9 |
| Tipologia prodotto Contributo in Collana |
| Elenco autori Federico Boschetti, Riccardo Del Gratta, Angelo Mario Del Grosso |
| Titolo The role of digital scholarly editors in the design of components for cooperative philology |
| Codice identificativo (ISBN) 978-90-8890-484-4 |
| anno pubblicazione 2017 |
| Altre informazioni Advances in Digital Scholarly Editing url: https://www.sidestone.com/books/advances-in-digital-scholarly-editing Note: ISBN 978-90-8890-483-7 (softcover) ISBN 978-90-8890-484-4 (hardcover) ISBN 978-90-8890-485-1 (PDF e-book) |

|  |
| --- |
| Nr 10 |
| Tipologia prodotto Articolo in Rivista |
| Elenco autori Mugelli G.; Boschetti F.; Del Gratta R.; Del Grosso A.M.; Khan F.; Taddei A. |
| Titolo A user-centred design to annotate ritual facts in ancient greek tragedies |
| Rivista Bulletin - University of London. Institute of Classical Studies |
| Codice identificativo (ISSN) 0076-0730 |
| anno pubblicazione 2016 |
| Indice di classificazione |
| Impact Factor rivista |
| ruolo svolto |
| numero citazioni |
| Altre informazioni doi: 10.1111/j.2041-5370.2016.12041.x url: http://www.scopus.com/record/display.url?eid=2-s2.0-85007489227&origin=inward abstract: Euporia is an annotation system developed with a user-centred approach for the study of ritual and religion in ancient Greek tragedy. Euporia adopts a domain specific language (DSL) and a lightweight web user interface in order to offer digital support to an anthropological study of ancient Greek tragedy that compares ritual as it is performed or described in Greek tragedy with ancient ritual as it can be reconstructed from literary, archaeological, and epigraphic sources. The case study discussed in this paper (Aesch. Ag 67-71) shows one of the main features of Euporia: the ability to annotate different readings and different interpretations of the text and their consequences in the reconstruction of ancient Greek ritual. |

|  |
| --- |
| Nr 11 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Berti Monica, Crane Gregory, Yousef Tariq, Bizzoni Yuri, Boschetti Federico, Del Gratta Riccardo |
| Titolo Ancient Greek WordNet meets the Dynamic Lexicon: the example of the fragments of the Greek Historians |
| Codice identificativo (ISBN) 978-606-714-239-6 |
| anno pubblicazione 2016 |
| abstract: The Ancient Greek WordNet (AGWN) and the Dynamic Lexicon (DL) are multilingual resources to study the lexicon of Ancient Greek texts and their translations. Both AGWN and DL are works in progress that need accuracy improvement and manual validation. After a detailed description of the current state of each work, this paper illustrates a methodology to cross AGWN and DL data, in order to mutually score the items of each resource according to the evidence provided by the other resource. The training data is based on the corpus of the Digital Fragmenta Historicorum Graecorum (DFHG), which includes ancient Greek texts with Latin translations. url: http://gwc2016.racai.ro/procedings.pdf abstract: The Ancient Greek WordNet (AGWN) and the Dynamic Lexicon (DL) are multilingual resources to study the lexicon of Ancient Greek texts and their translations. Both AGWN and DL are works in progress that need accuracy improvement and manual validation. After a detailed description of the current state of each work, this paper illustrates a methodology to cross AGWN and DL data, in order to mutually score the items of each resource according to the evidence provided by the other resource. The training data is based on the corpus of the Digital Fragmenta Historicorum Graecorum (DFHG), which includes ancient Greek texts with Latin translations. |

|  |
| --- |
| Nr 12 |
| Tipologia prodotto Contributo in Collana |
| Elenco autori Del Gratta R.; Boschetti F.; Del Grosso A.; Khan F.; Monachini M. |
| Titolo Cooperative philology on the way to web services: The case of the cophiwordnet platform |
| Codice identificativo (ISBN) 978-3-319-31468-6 |
| anno pubblicazione 2016 |
| Altre informazioni Worldwide Language Service Infrastructure: Second International Workshop, WLSI 2015, Kyoto, Japan, January 22-23, 2015. Revised Selected Papers url: http://www.scopus.com/inward/record.url?eid=2-s2.0-84961744725&partnerID=q2rCbXpz doi: 10.1007/978-3-319-31468-6\_13 ISSN: 0302-9743 |

|  |
| --- |
| Nr 13 |
| Tipologia prodotto Contributo in Collana |
| Elenco autori Francesca Frontini Riccardo Del Gratta Monica Monachini |
| Titolo GeoDomainWordNet: Linking the Geonames Ontology to WordNet |
| Codice identificativo (ISBN) 978-3-319-43808-5 |
| anno pubblicazione 2016 |
| Altre informazioni Human Language Technology. Challenges for Computer Science and Linguistics url: http://link.springer.com/chapter/10.1007/978-3-319-43808-5\_18 doi: 10.1007/978-3-319-43808-5 ISSN: 0302-9743 |

|  |
| --- |
| Nr 14 |
| Tipologia prodotto Contributo in Conferenza o Convegno (Senza ISBN) |
| Titolo Homeric Greek WordNet: costruire una risorsa lessico-semantica fra ricerca e didattica |
| Descrizione |
| Elenco autori Antonio Stanzione, Giulia Re, Gloria Mugelli, Andrea Taddei, Federico Boschetti, Riccardo Del Gratta |
| Ruolo svolto |
| anno pubblicazione 2016 |
| Altre informazioni In AIUCD 2016 - Book of Abstracts url: http://www.himeros.eu/aiucd2016/c40.pdf |

|  |
| --- |
| Nr 15 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Del Gratta R., Frontini F., Monachini M., Pardelli G., Russo I., Bartolini R., Khan F., Soria C., Calzolari N. |
| Titolo LREC as a Graph: People and Resources in a Network |
| Codice identificativo (ISBN) 978-2-9517408-9-1 |
| anno pubblicazione 2016 |
| abstract: This proposal describes a new way to visualise resources in the LREMap, a community-built repository of language resource descriptions and uses. The LREMap is represented as a force-directed graph, where resources, papers and authors are nodes. The analysis of the visual representation of the underlying graph is used to study how the community gathers around LRs and how LRs are used in research. url: http://www.lrec-conf.org/proceedings/lrec2016/index.html abstract: This proposal describes a new way to visualise resources in the LREMap, a community-built repository of language resource descriptions and uses. The LREMap is represented as a force-directed graph, where resources, papers and authors are nodes. The analysis of the visual representation of the underlying graph is used to study how the community gathers around LRs and how LRs are used in research. |

|  |
| --- |
| Nr 16 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Vladimir Popescu, Lin Liu, Riccardo Del Gratta, Khalid Choukri, Nicoletta Calzolari |
| Titolo New Developments in the LRE Map |
| Codice identificativo (ISBN) 978-2-9517408-9-1 |
| anno pubblicazione 2016 |
| abstract: In this paper we describe the new developments brought to LRE Map, especially in terms of the user interface of the Web application, of the searching of the information therein, and of the data model updates. Thus, users now have several new search facilities, such as faceted search and fuzzy textual search, they can now register, log in and store search bookmarks for further perusal. Moreover, the data model now includes the notion of paper and author, which allows for linking the resources to the scientific works. Also, users can now visualise author-provided field values and normalised values. The normalisation has been manual and enables a better grouping of the entries. Last but not least, provisions have been made towards linked open data (LOD) aspects, by exposing an RDF access point allowing to query on the authors, papers and resources. Finally, a complete technological overhaul of the whole application has been undertaken, especially in terms of the Web infrastructure and of the text search backend. url: http://www.lrec-conf.org/proceedings/lrec2016/pdf/1256\_Paper.pdf abstract: In this paper we describe the new developments brought to LRE Map, especially in terms of the user interface of the Web application, of the searching of the information therein, and of the data model updates. Thus, users now have several new search facilities, such as faceted search and fuzzy textual search, they can now register, log in and store search bookmarks for further perusal. Moreover, the data model now includes the notion of paper and author, which allows for linking the resources to the scientific works. Also, users can now visualise author-provided field values and normalised values. The normalisation has been manual and enables a better grouping of the entries. Last but not least, provisions have been made towards linked open data (LOD) aspects, by exposing an RDF access point allowing to query on the authors, papers and resources. Finally, a complete technological overhaul of the whole application has been undertaken, especially in terms of the Web infrastructure and of the text search backend. |

|  |
| --- |
| Nr 17 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Federico Boschetti, Riccardo Del Gratta, Francesca Frontini, Fahad Khan, Monica Monachini |
| Titolo (Re)thinking the BLARK for Ancient Greek |
| Codice identificativo (ISBN) 978-83-932640-8-7 |
| anno pubblicazione 2015 |
| abstract: The paper discusses the Basic LAnguage Resource Kit (BLARK) for Ancient Greek, measuring the BLARK matrix against what is actually available for this language, and assessing its applicability to ancient languages in general. In addition, the BLARK and the FLaReNet recommendations are used to define priorities in the sector in close collaboration between philologists and the broader LRT community.Proceedings of the Seventh Language and Technology Conference abstract: The paper discusses the Basic LAnguage Resource Kit (BLARK) for Ancient Greek, measuring the BLARK matrix against what is actually available for this language, and assessing its applicability to ancient languages in general. In addition, the BLARK and the FLaReNet recommendations are used to define priorities in the sector in close collaboration between philologists and the broader LRT community. |

|  |
| --- |
| Nr 18 |
| Tipologia prodotto Contributo in Conferenza o Convegno (Senza ISBN) |
| Titolo Collaborative Philology on the way to Web Services: the case of CoPhiWordnet |
| Descrizione |
| Elenco autori Federico Boschetti, Riccardo Del Gratta, Angelo Del Grosso, Monica Monachini, Harry Diakoff, and Ouafae Nahli |
| Ruolo svolto |
| anno pubblicazione 2015 |
| abstract: Starting from previous initiatives of the CoPhiLab, we show how they can be reinterpreted as Web Services, especially when they become part of a wider scenario: Web Services are used to make connections between lexicons, semantic resources and a  fine grained text management. Linked Open Data is chosen to be the paradigm used to link the di erent resources, but also as the modality of data presentation.In url: http://langrid.org/wlsi2015/program.html abstract: Starting from previous initiatives of the CoPhiLab, we show how they can be reinterpreted as Web Services, especially when they become part of a wider scenario: Web Services are used to make connections between lexicons, semantic resources and a  fine grained text management. Linked Open Data is chosen to be the paradigm used to link the di erent resources, but also as the modality of data presentation. |

|  |
| --- |
| Nr 19 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Pirrelli, Vito; Nahli, Ouafae; Boschetti, Federico; Del Gratta, Riccardo; Marzi, Claudia |
| Titolo Computational Linguistics and Language Physiology: Insights from Arabic NLP and Cooperative Editing |
| Codice identificativo (ISBN) 978-1-4503-3295-8 |
| anno pubblicazione 2015 |
| abstract: Computer processing of written Arabic raises a number of challenges to traditional parsing architectures on many levels of linguistic analysis. In this contribution, we review some of these core issues and the demands they make, to suggest different strategies to successfully tackle them. In the end, we assess these issues in connection with the behaviour of neuro-biologically inspired lexical architectures known as Temporal Self-Organising Maps. We show that, far from being language-specific problems, issues in Arabic processing can shed light on some fundamental characteristics of the human language processor, such as structure-based lexical recoding, concurrent, competitive activation of output candidates and dynamic selection of optimal solutions.Third AIUCD Annual Conference - Humanities and Their Methods in the Digital Ecosystem url: http://dl.acm.org/citation.cfm?id=2802612 abstract: Computer processing of written Arabic raises a number of challenges to traditional parsing architectures on many levels of linguistic analysis. In this contribution, we review some of these core issues and the demands they make, to suggest different strategies to successfully tackle them. In the end, we assess these issues in connection with the behaviour of neuro-biologically inspired lexical architectures known as Temporal Self-Organising Maps. We show that, far from being language-specific problems, issues in Arabic processing can shed light on some fundamental characteristics of the human language processor, such as structure-based lexical recoding, concurrent, competitive activation of output candidates and dynamic selection of optimal solutions. |

|  |
| --- |
| Nr 20 |
| Tipologia prodotto Articolo in Rivista |
| Elenco autori Del Gratta Riccardo, Francesca Frontini, Fahad Khan, Monica Monachini |
| Titolo Converting the PAROLE SIMPLE CLIPS Lexicon into RDF with lemon |
| Rivista Semantic web (Print) |
| Codice identificativo (ISSN) 1570-0844 |
| anno pubblicazione 2015 |
| Indice di classificazione |
| Impact Factor rivista |
| ruolo svolto |
| numero citazioni |
| Altre informazioni doi: 10.3233/SW-140168 url: http://www.semantic-web-journal.net/content/converting-parole-simple-clips-lexicon-rdf-lemon-0 abstract: This paper describes the publication and linking of (parts of) PAROLE SIMPLE CLIPS (PSC), a large scale Italian lexicon, to the Semantic Web and the Linked Data cloud using the lemon model. The main challenge of the conversion is discussed, namely the reconciliation between the PSC semantic structure which contains richly encoded semantic information, following the qualia structure of the Generative Lexicon theory and the lemon view of lexical sense as a reified pairing of a lexical item and a concept in an ontology. The result is two datasets: one consists of a list of lemon lexical entries with their lexical properties, relations and senses; the other consists of a list of OWL individuals representing the referents for the lexical senses. These OWL individuals are linked to each other by a set of semantic relations and mapped onto the SIMPLE OWL ontology of higher level semantic types. |

|  |
| --- |
| Nr 21 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Yuri Bizzoni, Federico Boschetti, Riccardo Del Gratta,Marianne Reboul |
| Titolo Enhancing the Accuracy of Ancient Greek WordNet by Multilingual Distributional Semantics |
| Codice identificativo (ISBN) 978-88-99200-62-6 |
| anno pubblicazione 2015 |
| abstract: We discuss a method to enhance the accuracy of a subset of the Ancient Greek WordNet based on the Homeric lexicon and the related conceptual network, by using multilingual semantic spaces built from aligned corpora. url: https://www.academia.edu/19413183/Enhancing\_the\_Accuracy\_of\_Ancient\_Greek\_WordNet\_by\_Multilingual\_Distributional\_Semantics abstract: We discuss a method to enhance the accuracy of a subset of the Ancient Greek WordNet based on the Homeric lexicon and the related conceptual network, by using multilingual semantic spaces built from aligned corpora. |

|  |
| --- |
| Nr 22 |
| Tipologia prodotto Contributo in Conferenza o Convegno (Senza ISBN) |
| Titolo Strumenti, Risorse e Linguistic Linked Open Data per le lingue antiche |
| Descrizione |
| Elenco autori Federico Boschetti, Riccardo Del Gratta, Francesca Frontini, Anas Fahad Khan, Monica Monachini |
| Ruolo svolto |
| anno pubblicazione 2015 |
| abstract: Strumenti e metodi dell'Informatica Umanistica hanno portato e portano ad una ridefinizione di processi teorici, metodologici e tecnici, fino a una vera e propria ri-concettualizzazione dei saperi nell'ambito dei beni culturali. L'Istituto di Linguistica Computazionale è attivo con varie iniziative sul fronte delle Digital Humanities per la creazione di strumenti e risorse linguistiche per il mondo classico. La direzione intrapresa si inserisce nel paradigma che si va consolidando nel settore delle tecnologie del linguaggio e che prevede la fruizione di servizi linguistici attraverso infrastrutture di ricerca, secondo un modello già operativo per le lingue moderne. Tale paradigma è in connessione con l'emergere degli standard e dei formati del web semantico per le tecnologie del linguaggio e per la pubblicazione di dati linguistici.In abstract: Strumenti e metodi dell\'Informatica Umanistica hanno portato e portano ad una ridefinizione di processi teorici, metodologici e tecnici, fino a una vera e propria ri-concettualizzazione dei saperi nell\'ambito dei beni culturali. L\'Istituto di Linguistica Computazionale è attivo con varie iniziative sul fronte delle Digital Humanities per la creazione di strumenti e risorse linguistiche per il mondo classico. La direzione intrapresa si inserisce nel paradigma che si va consolidando nel settore delle tecnologie del linguaggio e che prevede la fruizione di servizi linguistici attraverso infrastrutture di ricerca, secondo un modello già operativo per le lingue moderne. Tale paradigma è in connessione con l\'emergere degli standard e dei formati del web semantico per le tecnologie del linguaggio e per la pubblicazione di dati linguistici. |

|  |
| --- |
| Nr 23 |
| Tipologia prodotto Contributo in Conferenza o Convegno (Senza ISBN) |
| Titolo The role of digital scholarly editors in the design of components for cooperative philology |
| Descrizione |
| Elenco autori Angelo Mario Del Grosso, Riccardo Del Gratta, Federico Boschetti |
| Ruolo svolto |
| anno pubblicazione 2015 |
| abstract: This contribution is focused on the role of the digital scholarly editor in the continuous process of analysis, development and evaluation of libraries of components for cooperative philology. By following a general trend, in the domain of digital humanities developers are progressively shifting from the project-driven approach to the new community-driven paradigm. This shift is solicited by the increasing aggregation of scholars in communities of practice that are expressing common requirements and sharing best practices.  In most cases, service providers are responding to these needs by offering web services quickly developed by taking into account the specific functionality that they expose or, worse, by wrapping legacy code. Although a pipeline of web services devoted to linguistic analysis and collaborative annotation provides many advantages in terms of flexibility, we are concerned by the impact of the main drawbacks, in order to study alternative or complementary solutions for our domain.  Maintainability, performance and atomicity are the principal issues in which we are interested. In a chain of web services, the overall system depends by the status of the singles nodes and medium or small projects not always are able to grant the necessary level of redundancy or caching strategies. Performance is affected by the trade-off among challenging conditions (e.g. memory resources, computational overload, bandwidth). Atomicity influences the reusability and the extension of services (e.g. from many points of view, Latin metrical analysis is very similar to ancient Greek metrical analysis, but a web service that atomically provides the former could be totally unusable for the latter).  At the Cooperative Philology Lab (Institute of Computational Linguistics "A. Zampolli", CNR, Pisa) we try to address these issues by designing and developing a library of components for the domain of scholarly editing. A library can be installed locally or remotely and it provides multiple choices for maintenance and performance tuning. But above all a library of components provides the building blocks to shape local or remote services at the adequate level of atomicity, in order to ensure reusability and extendibility.  The role of the digital scholarly editors with which we have collaborated in pilot and funded projects at the CNR-ILC is crucial, because they are providing the necessary use cases that we are generalizing for the design of our library. During the workshop, we would like to stress the importance of a new generation of digital scholars that are not only creators of digital resources and consumers of computational tools or web infrastructures, but also actors in the analysis of requirements and in the evaluation of the libraries of components devoted to their activities.  References  Bozzi, "Computer-assisted scholarly editing of manuscript sources," in New publication cultures in the humanities: exploring the paradigm shift, Davidhazi, Ed. Amsterdam: Amsterdam University Press, 2014, pp. 99-115. [Online]. Available: http://www.oapen.org/record/515678  McGann, "From text to work: Digital tools and the emergence of the social text," Variants: The Journal of the European Society for Textual Scholarship, vol. 4, pp. 225-240, 2005.  Robinson, "Towards a scholarly editing system for the next decades," in Sanskrit Computational Linguistics, ser. Lecture Notes in Computer Science, G. Huet, A. Kulkarni, and P. Scharf, Eds. Springer Berlin Heidelberg, 2009, vol. 5402, pp. 346-357. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-00155-0 18  Robinson, "Towards a theory of digital editions," Variants, no. 10,105-131, 2013.  Siemens, M. Timney, C. Leitch, C. Koolen, A. Garnett et al., "Toward modeling the social edition: An approach to understanding the electronic scholarly edition in the context of new and emerging social media," Literary and Linguistic Computing, vol. 27, no. 4, pp. 445-461, 2012.In url: http://dixit.huygens.knaw.nl/?page\_id=138#boschetti abstract: This contribution is focused on the role of the digital scholarly editor in the continuous process of analysis, development and evaluation of libraries of components for cooperative philology. By following a general trend, in the domain of digital humanities developers are progressively shifting from the project-driven approach to the new community-driven paradigm. This shift is solicited by the increasing aggregation of scholars in communities of practice that are expressing common requirements and sharing best practices.  In most cases, service providers are responding to these needs by offering web services quickly developed by taking into account the specific functionality that they expose or, worse, by wrapping legacy code. Although a pipeline of web services devoted to linguistic analysis and collaborative annotation provides many advantages in terms of flexibility, we are concerned by the impact of the main drawbacks, in order to study alternative or complementary solutions for our domain.  Maintainability, performance and atomicity are the principal issues in which we are interested. In a chain of web services, the overall system depends by the status of the singles nodes and medium or small projects not always are able to grant the necessary level of redundancy or caching strategies. Performance is affected by the trade-off among challenging conditions (e.g. memory resources, computational overload, bandwidth). Atomicity influences the reusability and the extension of services (e.g. from many points of view, Latin metrical analysis is very similar to ancient Greek metrical analysis, but a web service that atomically provides the former could be totally unusable for the latter).  At the Cooperative Philology Lab (Institute of Computational Linguistics \"A. Zampolli\", CNR, Pisa) we try to address these issues by designing and developing a library of components for the domain of scholarly editing. A library can be installed locally or remotely and it provides multiple choices for maintenance and performance tuning. But above all a library of components provides the building blocks to shape local or remote services at the adequate level of atomicity, in order to ensure reusability and extendibility.  The role of the digital scholarly editors with which we have collaborated in pilot and funded projects at the CNR-ILC is crucial, because they are providing the necessary use cases that we are generalizing for the design of our library. During the workshop, we would like to stress the importance of a new generation of digital scholars that are not only creators of digital resources and consumers of computational tools or web infrastructures, but also actors in the analysis of requirements and in the evaluation of the libraries of components devoted to their activities.  References  Bozzi, \"Computer-assisted scholarly editing of manuscript sources,\" in New publication cultures in the humanities: exploring the paradigm shift, Davidhazi, Ed. Amsterdam: Amsterdam University Press, 2014, pp. 99-115. [Online]. Available: http://www.oapen.org/record/515678  McGann, \"From text to work: Digital tools and the emergence of the social text,\" Variants: The Journal of the European Society for Textual Scholarship, vol. 4, pp. 225-240, 2005.  Robinson, \"Towards a scholarly editing system for the next decades,\" in Sanskrit Computational Linguistics, ser. Lecture Notes in Computer Science, G. Huet, A. Kulkarni, and P. Scharf, Eds. Springer Berlin Heidelberg, 2009, vol. 5402, pp. 346-357. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-00155-0 18  Robinson, \"Towards a theory of digital editions,\" Variants, no. 10,105-131, 2013.  Siemens, M. Timney, C. Leitch, C. Koolen, A. Garnett et al., \"Toward modeling the social edition: An approach to understanding the electronic scholarly edition in the context of new and emerging social media,\" Literary and Linguistic Computing, vol. 27, no. 4, pp. 445-461, 2012. |

|  |
| --- |
| Nr 24 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Riccardo Del Gratta, Francesca Frontini, Monica Monachini, Gabriella Pardelli, Irene Russo, Roberto Bartolini, Sara Goggi, Fahad Khan, Valeria Quochi, Claudia Soria, Nicoletta Calzolari |
| Titolo Visualising Italian Language Resources: a Snapshot |
| Codice identificativo (ISBN) 978-88-99200-62-6 |
| anno pubblicazione 2015 |
| abstract: This paper aims to provide a first snapshot of Italian Language Resources (LRs) and their uses by the community, as documented by the papers presented at two different conferences, LREC2014 and CLiC-it 2014. The data of the former were drawn from the LOD version of the LRE Map, while those of the latter come from manually analyzing the proceedings. The results are presented in the form of visual graphs and confirm the initial hypothesis that Italian LRs require concrete actions to enhance their visibility.Proceedings of the Second Italian Conference on Computational Linguistics CLiC-it 2015 abstract: This paper aims to provide a first snapshot of Italian Language Resources (LRs) and their uses by the community, as documented by the papers presented at two different conferences, LREC2014 and CLiC-it 2014. The data of the former were drawn from the LOD version of the LRE Map, while those of the latter come from manually analyzing the proceedings. The results are presented in the form of visual graphs and confirm the initial hypothesis that Italian LRs require concrete actions to enhance their visibility. |

|  |
| --- |
| Nr 25 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Federico Boschetti, Riccardo Del Gratta, Marion Lamé |
| Titolo Computer Assisted Annotation of Themes and Motifs in Ancient Greek Epigrams: First Steps |
| Codice identificativo (ISBN) 978-8-86741-472-7 |
| anno pubblicazione 2014 |
| abstract: This paper aims at illustrating some tools to assist the manual annotation of themes and motifs in literary and epigraphic epigrams for the PRIN 2010/2011 Memorata Poetis Project.The First Italian Conference on Computational Linguistics - Proceedings url: http://clic.humnet.unipi.it/it/atti.html abstract: This paper aims at illustrating some tools to assist the manual annotation of themes and motifs in literary and epigraphic epigrams for the PRIN 2010/2011 Memorata Poetis Project. |

|  |
| --- |
| Nr 26 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Del Gratta Riccardo; Ouafae Nahli |
| Titolo Enhancing Arabic WordNet with the use on Princeton WordNet and a bilingual dictionary |
| Codice identificativo (ISBN) 978-1-4799-5978-5 |
| anno pubblicazione 2014 |
| abstract: This paper describes an heuristic-based approach to enhance existing WordNets with freely available bilingual resources. The approach has been applied to the Arabic WordNet using the AraMorph bilingual dictionary as bilingual resource, but its guidelines are quite general to be effectively applied to other languages. The English words extracted from the bilingual resource are checked against Princeton WordNet in order to quantify their coverage and to select only those words which share the same set of synsets. This strongly reduces the number of Arabic words of the pairs. These latter are then checked against the Arabic WordNet to make new words emerge and -possibly- add new synonyms.3rd International IEEE Colloquium on Information Science and Technology; From 20th to 22nd of October 2014 Tetuan - Chefchaouen Morocco url: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7016632 abstract: This paper describes an heuristic-based approach to enhance existing WordNets with freely available bilingual resources. The approach has been applied to the Arabic WordNet using the AraMorph bilingual dictionary as bilingual resource, but its guidelines are quite general to be effectively applied to other languages. The English words extracted from the bilingual resource are checked against Princeton WordNet in order to quantify their coverage and to select only those words which share the same set of synsets. This strongly reduces the number of Arabic words of the pairs. These latter are then checked against the Arabic WordNet to make new words emerge and -possibly- add new synonyms. |

|  |
| --- |
| Nr 27 |
| Tipologia prodotto Contributo in Conferenza o Convegno (Senza ISBN) |
| Titolo From text to cloud |
| Descrizione |
| Elenco autori Riccardo Del Gratta, Arsalane Zarghili, Ouafae Nahli, Angelo Del Grosso and Federico Boschetti |
| Ruolo svolto |
| anno pubblicazione 2014 |
| abstract: We describe a model to integrate tools for searching and editing texts with lexical­-semantic resources. We also suggest that this integration is necessary for students and researchers. Finally, we indicate (L)LOD as a formalism and strategy both for data organization and distribution.In url: http://www.dh.uni-leipzig.de/wo/workshop-december-2014/greek-and-latin-in-an-age-of-open-data-schedule/riccardo-del-gratta-et-al-from-text-to-cloud abstract: We describe a model to integrate tools for searching and editing texts with lexical­-semantic resources. We also suggest that this integration is necessary for students and researchers. Finally, we indicate (L)LOD as a formalism and strategy both for data organization and distribution. |

|  |
| --- |
| Nr 28 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Stelios Piperidis, Harris Papageorgiou, Christian Spurk, Georg Rehm, Khalid Choukri, Olivier Hamon, Nicoletta Calzolari, Riccardo del Gratta, Bernardo Magnini, Christian Girardi |
| Titolo META-SHARE: One year after |
| Codice identificativo (ISBN) 978-2-9517408-8-4 |
| anno pubblicazione 2014 |
| abstract: This paper presents META-SHARE (www.meta-share.eu), an open language resource infrastructure, and its usage since its Europe-wide deployment in early 2013. META-SHARE is a network of repositories that store language resources (data, tools and processing services) documented with high-quality metadata, aggregated in central inventories allowing for uniform search and access. META-SHARE was developed by META-NET (www.meta-net.eu) and aims to serve as an important component of a language technology marketplace for researchers, developers, professionals and industrial players, catering for the full development cycle of language technology, from research through to innovative products and services. The observed usage in its initial steps, the steadily increasing number of network nodes, resources, users, queries, views and downloads are all encouraging and considered as supportive of the choices made so far. In tandem, take-up activities like direct linking and processing of datasets by language processing services as well as metadata transformation to RDF are expected to open new avenues for data and resources linking and boost the organic growth of the infrastructure while facilitating language technology deployment by much wider research communities and industrial sectors.Proceedings of the Ninth International Conference on Language Resources and Evaluation url: http://www.lrec-conf.org/proceedings/lrec2014/index.html abstract: This paper presents META-SHARE (www.meta-share.eu), an open language resource infrastructure, and its usage since its Europe-wide deployment in early 2013. META-SHARE is a network of repositories that store language resources (data, tools and processing services) documented with high-quality metadata, aggregated in central inventories allowing for uniform search and access. META-SHARE was developed by META-NET (www.meta-net.eu) and aims to serve as an important component of a language technology marketplace for researchers, developers, professionals and industrial players, catering for the full development cycle of language technology, from research through to innovative products and services. The observed usage in its initial steps, the steadily increasing number of network nodes, resources, users, queries, views and downloads are all encouraging and considered as supportive of the choices made so far. In tandem, take-up activities like direct linking and processing of datasets by language processing services as well as metadata transformation to RDF are expected to open new avenues for data and resources linking and boost the organic growth of the infrastructure while facilitating language technology deployment by much wider research communities and industrial sectors. |

|  |
| --- |
| Nr 29 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Del Gratta Riccardo, Pardelli Gabriella, Sara Goggi |
| Titolo The LRE Map disclosed |
| Codice identificativo (ISBN) 978-2-9517408-8-4 |
| anno pubblicazione 2014 |
| abstract: This paper describes a serialization of the LRE Map database according to the RDF model. Due to the peculiar nature of the LRE Map, many ontologies are necessary to model the map in RDF, including newly created and reused ontologies. The importance of having the LRE Map in RDF and its connections to other open resources is also addressed.Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC\'14) url: http://www.lrec-conf.org/proceedings/lrec2014/index.html abstract: This paper describes a serialization of the LRE Map database according to the RDF model. Due to the peculiar nature of the LRE Map, many ontologies are necessary to model the map in RDF, including newly created and reused ontologies. The importance of having the LRE Map in RDF and its connections to other open resources is also addressed. |

|  |
| --- |
| Nr 30 |
| Tipologia prodotto Contributo in Conferenza o Convegno (Senza ISBN) |
| Titolo The LREMap for Under-Resourced Languages |
| Descrizione |
| Elenco autori Riccardo Del Gratta, Francesca Frontini, Fahad Khan, Joseph Mariani, Claudia Soria |
| Ruolo svolto |
| anno pubblicazione 2014 |
| abstract: A complete picture of currently available language resources and technologies for the under-resourced languages of Europe is still lacking. Yet this would help policy makers, researchers and developers enormously in planning a roadmap for providing all languages with the necessary instruments to act as fully equipped languages in the digital era. In this paper we introduce the LRE Map and show its utility for documenting available language resources and technologies for under-resourced languages. The importance of the serialization of the LREMap into (L)LOD along with the possibility of its connection to a wider world is also introduced.In Proceedings of the Workshop on Collaboration and Computing for Under-Resourced Languages in the Linked Open Data Era (CCURL 2014) url: http://www.lrec-conf.org/proceedings/lrec2014/index.html abstract: A complete picture of currently available language resources and technologies for the under-resourced languages of Europe is still lacking. Yet this would help policy makers, researchers and developers enormously in planning a roadmap for providing all languages with the necessary instruments to act as fully equipped languages in the digital era. In this paper we introduce the LRE Map and show its utility for documenting available language resources and technologies for under-resourced languages. The importance of the serialization of the LREMap into (L)LOD along with the possibility of its connection to a wider world is also introduced. |

|  |
| --- |
| Nr 31 |
| Tipologia prodotto Contributo in Conferenza o Convegno (Senza ISBN) |
| Titolo The Making of Ancient Greek WordNet |
| Descrizione |
| Elenco autori Bizzoni Yuri, Federico Boschetti, Harry Diakoff, Riccardo Del Gratta, Monica Monachini, Gregory Crane |
| Ruolo svolto |
| anno pubblicazione 2014 |
| abstract: This paper describes the process of creation and review of a new lexico-semantic resource for the classical studies: AncientGreekWord- Net. The candidate sets of synonyms (synsets) are extracted from Greek-English dictionaries, on the assumption that Greek words translated by the same English word or phrase have a high probability of being synonyms or at least semantically closely related. The process of validation and the web interface developed to edit and query the resource are described in detail. The lexical coverage of Ancient Greek WordNet is illustrated and the accuracy is evaluated. Finally, scenarios for exploiting the resource are discussed.In url: http://www.lrec-conf.org/proceedings/lrec2014/index.html abstract: This paper describes the process of creation and review of a new lexico-semantic resource for the classical studies: AncientGreekWord- Net. The candidate sets of synonyms (synsets) are extracted from Greek-English dictionaries, on the assumption that Greek words translated by the same English word or phrase have a high probability of being synonyms or at least semantically closely related. The process of validation and the web interface developed to edit and query the resource are described in detail. The lexical coverage of Ancient Greek WordNet is illustrated and the accuracy is evaluated. Finally, scenarios for exploiting the resource are discussed. |

|  |
| --- |
| Nr 32 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Fahad Khan, Francesca Frontini, Riccardo Del Gratta, Monica Monachini, Valeria Quochi |
| Titolo Generative Lexicon Theory and Linguistic Linked Open Data |
| Codice identificativo (ISBN) 978-1-937284-98-5 |
| anno pubblicazione 2013 |
| abstract: In this paper we look at how Generative Lexicon theory can assist in providing a more thorough definition of word senses as links between items in a RDF-based lexicon and concepts in an ontology. We focus on the definition of lexical sense in lemon and show its limitations before defining a new model based on lemon and which we term lemonGL. This new model is an initial attempt at providing a way of structuring lexico-ontological resources as linked data in such a way as to allow a rich representation of word meaning (following the GL theory) while at the same time (attempting to) re-main faithful to the separation between the lexicon and the ontology as recommended by the lemon model.Proceedings of the 6th International Conference on Generative Approaches to the Lexicon. Generative Lexicon and Distributional Semantics. url: http://aclweb.org/anthology/W13-54 abstract: In this paper we look at how Generative Lexicon theory can assist in providing a more thorough definition of word senses as links between items in a RDF-based lexicon and concepts in an ontology. We focus on the definition of lexical sense in lemon and show its limitations before defining a new model based on lemon and which we term lemonGL. This new model is an initial attempt at providing a way of structuring lexico-ontological resources as linked data in such a way as to allow a rich representation of word meaning (following the GL theory) while at the same time (attempting to) re-main faithful to the separation between the lexicon and the ontology as recommended by the lemon model. |

|  |
| --- |
| Nr 33 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Francesca Frontini, Riccardo Del Gratta, Monica Monachini. |
| Titolo Linking the Geonames ontology to WordNet |
| Codice identificativo (ISBN) 978-2-9517408-8-4 |
| anno pubblicazione 2013 |
| abstract: This paper illustrates the transformation of the GeoNames ontology concepts, with their English labels and glosses, into a GeoDomain WordNet-like resource in English, its translation into Italian, and its linking to the existing generic WordNets of both languages.Human Language Technologies as a Challenge for Computer Science and Linguistics. Proceedings, 6th Language \& Technology Conference, December 7-9, 2013, Poznañ, Poland url: http://hnk.ffzg.hr/bibl/ltc2013/book/papers/OWN-2.pdf abstract: This paper illustrates the transformation of the GeoNames ontology concepts, with their English labels and glosses, into a GeoDomain WordNet-like resource in English, its translation into Italian, and its linking to the existing generic WordNets of both languages. |

|  |
| --- |
| Nr 34 |
| Tipologia prodotto Deliverable di Progetto |
| Titolo D3.4 Third version (v4) of the integrated platform and documentation |
| Descrizione |
| Elenco autori Poch, Marc; Hamon, Olivier; Quochi, Valeria; Del Gratta, Riccardo; Toral, Antonio; Thurmair, Gregor; Prokopidis, Prokopis; Bel, Nuria |
| Ruolo svolto |
| anno pubblicazione 2012 |
| Altre informazioni |

|  |
| --- |
| Nr 35 |
| Tipologia prodotto Deliverable di Progetto |
| Titolo D6.3 Monolingual lexica for English, Spanish and Italian tuned for a particular domain (LAB and ENV) |
| Descrizione |
| Elenco autori Rimell, Laura; Bel, Nuria; Padrò, Muntsa; Frontini, Francesca; Monachini, Monica; Quochi, Valeria; Del Gratta, Riccardo |
| Ruolo svolto |
| anno pubblicazione 2012 |
| Altre informazioni |

|  |
| --- |
| Nr 36 |
| Tipologia prodotto Deliverable di Progetto |
| Titolo D6.5 Merged dictionaries |
| Descrizione |
| Elenco autori Rimell, Laura; Bel, Núria; Padró, Muntsa; Frontini, Francesca; Monachini, Monica; Quochi, Valeria; Del Gratta, Riccardo |
| Ruolo svolto |
| anno pubblicazione 2012 |
| Altre informazioni |

|  |
| --- |
| Nr 37 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Riccardo Del Gratta, Francesca Frontini, Monica Monachini, Valeria Quochi, Francesco Rubino, Matteo Abrate, Angelica Lo Duca |
| Titolo L-LEME: an Automatic Lexical Merger based on the LMF Standard |
| Codice identificativo (ISBN) 978-2-9517408-7-7 |
| anno pubblicazione 2012 |
| abstract: The present paper describes LMF LExical MErger (L-LEME), an architecture to combine two lexicons in order to obtain new resource(s). L-LEME relies on standards, thus exploiting the benefits of the ISO Lexical Markup Framework (LMF) to ensure interoperability. L-LEME is meant to be dynamic and heavily adaptable: it allows the users to configure it to meet their specific needs. The L-LEME architecture is composed of two main modules: the Mapper, which takes in input two lexicons A and B and a set of user-defined rules and instructions to guide the mapping process (Directives D) and gives in output all matching entries. The algorithm also calculates a cosine similarity score. The Builder takes in input the previous results, a set of Directives D1 and produces a new LMF lexicon C. The Directives allow the user to define its own building rules and different merging scenarios. L-LEME is applied to a specific concrete task within the PANACEA project, namely the merging of two Italian SubCategorization Frame (SCF) lexicons. The experiment is interesting in that A and B have different philosophies behind, being A built by human introspection and B automatically extracted. Ultimately, L-LEME has interesting repercussions in many language technology applicationsProceedings of the LREC 2012 Workshop on Language Resource Merging abstract: The present paper describes LMF LExical MErger (L-LEME), an architecture to combine two lexicons in order to obtain new resource(s). L-LEME relies on standards, thus exploiting the benefits of the ISO Lexical Markup Framework (LMF) to ensure interoperability. L-LEME is meant to be dynamic and heavily adaptable: it allows the users to configure it to meet their specific needs. The L-LEME architecture is composed of two main modules: the Mapper, which takes in input two lexicons A and B and a set of user-defined rules and instructions to guide the mapping process (Directives D) and gives in output all matching entries. The algorithm also calculates a cosine similarity score. The Builder takes in input the previous results, a set of Directives D1 and produces a new LMF lexicon C. The Directives allow the user to define its own building rules and different merging scenarios. L-LEME is applied to a specific concrete task within the PANACEA project, namely the merging of two Italian SubCategorization Frame (SCF) lexicons. The experiment is interesting in that A and B have different philosophies behind, being A built by human introspection and B automatically extracted. Ultimately, L-LEME has interesting repercussions in many language technology applications |

|  |
| --- |
| Nr 38 |
| Tipologia prodotto Contributo in Conferenza o Convegno (Senza ISBN) |
| Titolo The Language Library: supporting community effort for collective resource production |
| Descrizione |
| Elenco autori Del Gratta, Riccardo; Frontini, Francesca; Rubino, Francesco; Russo, Irene; Calzolari, Nicoletta |
| Ruolo svolto |
| anno pubblicazione 2012 |
| abstract: Relations among phenomena at different linguistic levels are at the essence of language properties but today we focus mostly on one specific linguistic layer at a time, without (having the possibility of) paying attention to the relations among the different layers. At the same time our efforts are too much scattered without much possibility of exploiting other people's achievements. To address the complexities hidden in multilayer interrelations even small amounts of processed data can be useful, improving the performance of complex systems. Exploiting the current trend towards sharing we want to initiate a collective movement that works towards creating synergies and harmonisation among different annotation efforts that are now dispersed. In this paper we present the general architecture of the Language Library, an initiative which is conceived as a facility for gathering and making available through simple functionalities the linguistic knowledge the field is able to produce, putting in place new ways of collaboration within the LRT community. In order to reach this goal, a first population round of the Language Library has started around a core of parallel/comparable texts that have been annotated by several contributors submitting a paper for LREC2012. The Language Library has also an ancillary aim related to language documentation and archiving and it is conceived as a theory-neutral space which allows for several language processing philosophies to coexist.In abstract: Relations among phenomena at different linguistic levels are at the essence of language properties but today we focus mostly on one specific linguistic layer at a time, without (having the possibility of) paying attention to the relations among the different layers. At the same time our efforts are too much scattered without much possibility of exploiting other people\'s achievements. To address the complexities hidden in multilayer interrelations even small amounts of processed data can be useful, improving the performance of complex systems. Exploiting the current trend towards sharing we want to initiate a collective movement that works towards creating synergies and harmonisation among different annotation efforts that are now dispersed. In this paper we present the general architecture of the Language Library, an initiative which is conceived as a facility for gathering and making available through simple functionalities the linguistic knowledge the field is able to produce, putting in place new ways of collaboration within the LRT community. In order to reach this goal, a first population round of the Language Library has started around a core of parallel/comparable texts that have been annotated by several contributors submitting a paper for LREC2012. The Language Library has also an ancillary aim related to language documentation and archiving and it is conceived as a theory-neutral space which allows for several language processing philosophies to coexist. |

|  |
| --- |
| Nr 39 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Calzolari, Nicoletta [1]; Del Gratta, Riccardo [1]; Francopoulo, Gil [2]; Mariani, Joseph [2]; Rubino, Francesco [1]; Russo, Irene [1]; Soria, Claudia [1] |
| Titolo The LRE Map. Harmonising Community Descriptions of Resources |
| Codice identificativo (ISBN) 978-2-9517408-7-7 |
| anno pubblicazione 2012 |
| abstract: Accurate and reliable documentation of Language Resources is an undisputable need: documentation is the gateway to discovery of Language Resources, a necessary step towards promoting the data economy. Language resources that are not documented virtually do not exist: for this reason every initiative able to collect and harmonise metadata about resources represents a valuable opportunity for the NLP community. In this paper we describe the LRE Map, reporting statistics on resources associated with LREC2012 papers and providing comparisons with LREC2010 data. The LRE Map, jointly launched by FLaReNet and ELRA in conjunction with the LREC 2010 conference, is an instrument for enhancing availability of information about resources, either new or already existing ones, reinforcing and facilitating the use of standards in the community. The LRE Map web interface provides the possibility of searching according to a fixed set of metadata and to view the details of extracted resources. The LRE Map is continuing to collect bottom-up input about resources from authors of other conferences through standard submission process. This will help broadening the notion of "language resources" and attract to the field neighboring disciplines that so far have been only marginally involved by the standard notion of language resources.Proceedings of LREC\'12 - The Eight International Conference on Language Resources and Evaluation url: http://www.lrec-conf.org/proceedings/lrec2012/index.html abstract: Accurate and reliable documentation of Language Resources is an undisputable need: documentation is the gateway to discovery of Language Resources, a necessary step towards promoting the data economy. Language resources that are not documented virtually do not exist: for this reason every initiative able to collect and harmonise metadata about resources represents a valuable opportunity for the NLP community. In this paper we describe the LRE Map, reporting statistics on resources associated with LREC2012 papers and providing comparisons with LREC2010 data. The LRE Map, jointly launched by FLaReNet and ELRA in conjunction with the LREC 2010 conference, is an instrument for enhancing availability of information about resources, either new or already existing ones, reinforcing and facilitating the use of standards in the community. The LRE Map web interface provides the possibility of searching according to a fixed set of metadata and to view the details of extracted resources. The LRE Map is continuing to collect bottom-up input about resources from authors of other conferences through standard submission process. This will help broadening the notion of \"language resources\" and attract to the field neighboring disciplines that so far have been only marginally involved by the standard notion of language resources. |

|  |
| --- |
| Nr 40 |
| Tipologia prodotto Articolo in Rivista |
| Elenco autori Paul Thompson, John McNaught, Simonetta Montemagni, Nicoletta Calzolari, Riccardo del Gratta, Vivian Lee, Simone Marchi, Monica Monachini, Piotr Pezik, Valeria Quochi, CJ Rupp, Yutaka Sasaki, Giulia Venturi, Dietrich Rebholz-Schuhmann, Sophia Ananiadou |
| Titolo The BioLexicon: a large-scale terminological resource for biomedical text mining |
| Rivista BMC bioinformatics |
| Codice identificativo (ISSN) 1471-2105 |
| anno pubblicazione 2011 |
| Indice di classificazione |
| Impact Factor rivista |
| ruolo svolto |
| numero citazioni |
| Altre informazioni doi: 10.1186/1471-2105-12-397 url: http://www.biomedcentral.com/1471-2105/12/397 abstract: Background Due to the rapidly expanding body of biomedical literature, biologists require increasingly sophisticated and efficient systems to help them to search for relevant information. Such systems should account for the multiple written variants used to represent biomedical concepts, and allow the user to search for specific pieces of knowledge (or events) involving these concepts, e.g., protein-protein interactions. Such functionality requires access to detailed information about words used in the biomedical literature. Existing databases and ontologies often have a specific focus and are oriented towards human use. Consequently, biological knowledge is dispersed amongst many resources, which often do not attempt to account for the large and frequently changing set of variants that appear in the literature. Additionally, such resources typically do not provide information about how terms relate to each other in texts to describe events. Results This article provides an overview of the design, construction and evaluation of a large-scale lexical and conceptual resource for the biomedical domain, the BioLexicon. The resource can be exploited by text mining tools at several levels, e.g., part-of-speech tagging, recognition of biomedical entities, and the extraction of events in which they are involved. As such, the BioLexicon must account for real usage of words in biomedical texts. In particular, the BioLexicon gathers together different types of terms from several existing data resources into a single, unified repository, and augments them with new term variants automatically extracted from biomedical literature. Extraction of events is facilitated through the inclusion of biologically pertinent verbs (around which events are typically organized) together with information about typical patterns of grammatical and semantic behaviour, which are acquired from domain-specific texts. In order to foster interoperability, the BioLexicon is modelled using the Lexical Markup Framework, an ISO standard. Conclusions The BioLexicon contains over 2.2 M lexical entries and over 1.8 M terminological variants, as well as over 3.3 M semantic relations, including over 2 M synonymy relations. Its exploitation can benefit both application developers and users. We demonstrate some such benefits by describing integration of the resource into a number of different tools, and evaluating improvements in performance that this can bring. |

|  |
| --- |
| Nr 41 |
| Tipologia prodotto Deliverable di Progetto |
| Titolo The FLaReNet Databook |
| Descrizione |
| Elenco autori Arranz V., Bel N., Budin G., Caselli T., Choukri K., Del Gratta R., Frontini F., Goggi S., Monachini M., Quochi V., Rubino F., Russo I. et alii |
| Ruolo svolto |
| anno pubblicazione 2011 |
| Altre informazioni |

|  |
| --- |
| Nr 42 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Nicoletta Calzolari, Riccardo Del Gratta, Francesca Frontini, Irene Russo |
| Titolo The Language Library: Many Layers, More Knowledge |
| Codice identificativo (ISBN) 978-974-466-564-5 |
| anno pubblicazione 2011 |
| abstract: In this paper we outline the general concept of the Language Library, a new initiative that has the purpose of building a huge archive of structured colletion of linguistic information. The Language Library is conceived as a community built repository and as an environment that allows language specialists to share multidimensional and multi-level annotated/processed resources. The first steps towards its implementation are briefly sketched. abstract: In this paper we outline the general concept of the Language Library, a new initiative that has the purpose of building a huge archive of structured colletion of linguistic information. The Language Library is conceived as a community built repository and as an environment that allows language specialists to share multidimensional and multi-level annotated/processed resources. The first steps towards its implementation are briefly sketched. |

|  |
| --- |
| Nr 43 |
| Tipologia prodotto Contributo in Conferenza o Convegno (Senza ISBN) |
| Titolo A bilingual dictionary Mexican Sign Language-Spanish/Spanish-Mexican Sign Language |
| Descrizione |
| Elenco autori Hawayek A.; Del Gratta R.; Cappelli G. |
| Ruolo svolto |
| anno pubblicazione 2010 |
| Altre informazioni In |

|  |
| --- |
| Nr 44 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Del Gratta R.; D\'Onofrio L.; Bartolini R.; Caselli T.; Enea A.; Monachini M.; Quochi V.; Soria C.; Toral A.; Calzolari N. |
| Titolo A Web-based Architecture for Interoperability of Lexical Resources |
| Codice identificativo (ISBN) 978-962-442-323-5 |
| anno pubblicazione 2010 |
| abstract: In this paper we present aWeb Service Architecture for managing high level interoperability of Language Resources (LRs) by means of a Service Oriented Architecture (SOA) and the use of ISO standards, such as ISO LMF. We propose a layered architecture which separates the management of legacy resources (data collection) from data aggregation (workflow) and data access (user requests). We provide a case study to demonstrate how the proposed architecture is capable of managing data exchange among different lexical services in a coherent way and show how the use of a lexical standard becomes of primary importance when a protocol of interoperability is defined.2nd International Conference on Global Interoperability for Language Resources , ICGL 2010 url: http://weblab.iit.cnr.it/kyoto/www2.let.vu.nl/twiki/pub/Kyoto/Publications/icgl2010\_DOnofrioetal.pdf abstract: In this paper we present aWeb Service Architecture for managing high level interoperability of Language Resources (LRs) by means of a Service Oriented Architecture (SOA) and the use of ISO standards, such as ISO LMF. We propose a layered architecture which separates the management of legacy resources (data collection) from data aggregation (workflow) and data access (user requests). We provide a case study to demonstrate how the proposed architecture is capable of managing data exchange among different lexical services in a coherent way and show how the use of a lexical standard becomes of primary importance when a protocol of interoperability is defined. |

|  |
| --- |
| Nr 45 |
| Tipologia prodotto Articolo in Rivista |
| Elenco autori Calzolari N.; Soria C.; Del Gratta R. |
| Titolo The LREC 2010 Map of Language Resources and Tools |
| Rivista |
| Rivista |
| anno pubblicazione 2010 |
| Indice di classificazione |
| Impact Factor rivista |
| ruolo svolto |
| numero citazioni |
| Altre informazioni |

|  |
| --- |
| Nr 46 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Calzolari N.; Soria C.; Del Gratta R.; Goggi S.; Quochi V.; Russo I.; Choukri K.; Mariani J.; Piperidis S. |
| Titolo The LREC Map of Language Resources and Technologies |
| Codice identificativo (ISBN) 2-9517408-6-7 |
| anno pubblicazione 2010 |
| abstract: In this paper we present the LREC Map of Language Resources (data and tools), an innovative feature introduced in conjunction with the LREC 2010 Conference. The purpose of the Map is to shed light on the vast amount of resources that represent the background of the research presented at LREC, in the attempt to fill in a gap in the community knowledge about the resources that are used or created worldwide. It also aims at a change of culture in the field, actively engaging each researcher in the documentation task about resources. The Map has been developed on the basis of the information provided by LREC authors during the submission of papers to the LREC 2010 conference and the LREC workshops, and contains information about almost 2000 resources. The paper illustrates the motivation behind this initiative, its main characteristics, its relevance and future impact in the field, the metadata used to describe the resources, and finally presents some of the most relevant findings.LREC\'10 - Seventh International Conference on Language Resources and Evaluation. Proceedings url: http://www.lrec-conf.org/proceedings/lrec2010/index.html abstract: In this paper we present the LREC Map of Language Resources (data and tools), an innovative feature introduced in conjunction with the LREC 2010 Conference. The purpose of the Map is to shed light on the vast amount of resources that represent the background of the research presented at LREC, in the attempt to fill in a gap in the community knowledge about the resources that are used or created worldwide. It also aims at a change of culture in the field, actively engaging each researcher in the documentation task about resources. The Map has been developed on the basis of the information provided by LREC authors during the submission of papers to the LREC 2010 conference and the LREC workshops, and contains information about almost 2000 resources. The paper illustrates the motivation behind this initiative, its main characteristics, its relevance and future impact in the field, the metadata used to describe the resources, and finally presents some of the most relevant findings. |

|  |
| --- |
| Nr 47 |
| Tipologia prodotto Articolo in Rivista |
| Elenco autori Quochi V.; Del Gratta R.; Sassolini E.; Bartolini R.; Monachini M.; Calzolari N. |
| Titolo A Standard Lexical-Terminological Resource for the Bio Domain |
| Rivista Lecture notes in computer science |
| Codice identificativo (ISSN) 0302-9743 |
| anno pubblicazione 2009 |
| Indice di classificazione |
| Impact Factor rivista |
| ruolo svolto |
| numero citazioni |
| Altre informazioni doi: 10.1007/978-3-642-04235-5\_28 abstract: The present paper describes a large-scale lexical resource for the biology domain designed both for human and for machine use. This lexicon aims at semantic interoperability and extendability, through the adoption of ISO-LMF standard for lexical representation and through a granular and distributed encoding of relevant information. The first part of this contribution focuses on three aspects of the model that are of particular interest to the biology community: the treatment of term variants, the representation on bio events and the alignment with a domain ontology. The second part of the paper describes the physical implementation of the model: a relational database equipped with a set of automatic uploading procedures. Peculiarity of the BioLexicon is that it combines features of both terminologies and lexicons. A set verbs relevant for the domain is also represented with full details on their syntactic and semantic argument structure. |

|  |
| --- |
| Nr 48 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Quochi V.; Monachini M.; Del Gratta R.; Calzolari N. |
| Titolo A lexicon for biology and bioinformatics: the BOOTStrep experience |
| Codice identificativo (ISBN) 2-9517408-4-0 |
| anno pubblicazione 2008 |
| Altre informazioni url: http://www.lrec-conf.org/proceedings/lrec2008/pdf/576\_paper.pdf |

|  |
| --- |
| Nr 49 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Rebholz-Schuhmann D.; Pezik P.; Lee V.; Del Gratta R.; Kim J.; Sasaki Y.; McNaught J.; Montemagni S.; Monachini M.; Calzolari N.; Ananiadou S. |
| Titolo BioLexicon: Towards a reference terminological resource in the biomedical domain |
| Codice identificativo (ISBN) 978-1-61567-371-1 |
| anno pubblicazione 2008 |
| abstract: The BioLexicon is a publicly available large-scale terminological resource which brings together potential terms from several resources representing selected semantic types (genes, proteins, chemicals, species, enzymes, selected ontological terms). The schema of the BioLexicon enables improved resolution of term ambiguity and follows lexical standards for terminological resources. abstract: The BioLexicon is a publicly available large-scale terminological resource which brings together potential terms from several resources representing selected semantic types (genes, proteins, chemicals, species, enzymes, selected ontological terms). The schema of the BioLexicon enables improved resolution of term ambiguity and follows lexical standards for terminological resources. |

|  |
| --- |
| Nr 50 |
| Tipologia prodotto Contributo in Conferenza o Convegno (Senza ISBN) |
| Titolo Lexically-based Ontologies and Ontologically Based Lexicons |
| Descrizione |
| Elenco autori Toral Ruiz A.; Quochi V.; Del Gratta R.; Monachini M.; Soria C.; Calzolari N. |
| Ruolo svolto |
| anno pubblicazione 2008 |
| abstract: This paper deals with the relations between ontologies and lexicons. We study the role of these two components and their evolution during the last years in the field of Computational Linguistics. Subsequently, we survey the current lines of research at ILC-CNR which tackle this topic. They involve (I) the reuse of already existing Lexical Resources to derive formal ontologies, (II) the conversion and combination of terminologies into rich and formal Lexical Resources and (III) the use of formal ontologies as the backbone of multilingual Lexical Resources.In abstract: This paper deals with the relations between ontologies and lexicons. We study the role of these two components and their evolution during the last years in the field of Computational Linguistics. Subsequently, we survey the current lines of research at ILC-CNR which tackle this topic. They involve (I) the reuse of already existing Lexical Resources to derive formal ontologies, (II) the conversion and combination of terminologies into rich and formal Lexical Resources and (III) the use of formal ontologies as the backbone of multilingual Lexical Resources. |

|  |
| --- |
| Nr 51 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Del Gratta, Riccardo; Ruimy, Nilda; Toral, Antonio |
| Titolo Simple-Clips ongoing research: more information with less data by implementing inheritance |
| Codice identificativo (ISBN) 2-9517408-4-0 |
| anno pubblicazione 2008 |
| abstract: This paper presents the application of inheritance to the formal taxonomy (is-a) of a semantically rich Lexical Resource (LR) based on the Generative Lexicon theory, SIMPLE-CLIPS. The aim is to lighten the representation of its semantic layer by reducing the number of encoded relations. A prediction calculation on the impact of introducing inheritance as regards space occupancy is carried out, which yields a significant space reduction of 22\%. This is corroborated by its actual application that reduces the number of explicitly encoded relations in this lexicon by 18.4\%. Later on, we study the issues that inheritance poses to the Lexical Resources and discuss sensitive solutions, illustrated by examples, to tackle each of them. Finally, we present a discussion on the application of inheritance, from which two advantages arise: consistency enhancement and inference capabilities. abstract: This paper presents the application of inheritance to the formal taxonomy (is-a) of a semantically rich Lexical Resource (LR) based on the Generative Lexicon theory, SIMPLE-CLIPS. The aim is to lighten the representation of its semantic layer by reducing the number of encoded relations. A prediction calculation on the impact of introducing inheritance as regards space occupancy is carried out, which yields a significant space reduction of 22\%. This is corroborated by its actual application that reduces the number of explicitly encoded relations in this lexicon by 18.4\%. Later on, we study the issues that inheritance poses to the Lexical Resources and discuss sensitive solutions, illustrated by examples, to tackle each of them. Finally, we present a discussion on the application of inheritance, from which two advantages arise: consistency enhancement and inference capabilities. |

|  |
| --- |
| Nr 52 |
| Tipologia prodotto Contributo in Conferenza o Convegno (Senza ISBN) |
| Titolo TimeML: An ontological mapping onto UIMA Type Systems |
| Descrizione |
| Elenco autori Del Gratta R.; Caselli T.; Calzolari N.; Ruimy N. |
| Ruolo svolto |
| anno pubblicazione 2008 |
| Altre informazioni In |

|  |
| --- |
| Nr 53 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Del Gratta R.; Bartolini R.; Caselli T.; Monachini M.; Soria C.; Calzolari N. |
| Titolo UFRA: a UIMA-based Approach to Federated Language Resource Architecture |
| Codice identificativo (ISBN) 2-9517408-4-0 |
| anno pubblicazione 2008 |
| abstract: In this paper we address the issue of developing an interoperable infrastructure for language resources and technologies. In our approach, called UFRA, we extend the Federate Database Architecture System adding typical functionalities caming from UIMA. In this way, we capitalize the advantages of a federated architecture, such as autonomy, heterogeneity and distribution of components, monitored by a central authority responsible for checking both the integration of components and user rights on performing different tasks. We use the UIMA approach to manage and define one common front-end, enabling users and clients to query, retrieve and use language resources and technologies. The purpose of this paper is to show how UIMA leads from a Federated Database Architecture to a Federated Resource Architecture, adding to a registry of available components both static resources such as lexicons and corpora and dynamic ones such as tools and general purpose language technologies. At the end of the paper, we present a case-study that adopts this framework to integrate the SIMPLE lexicon and TIMEML annotation guidelines to tag natural language texts. url: http://www.lrec-conf.org/proceedings/lrec2008/pdf/656\_paper.pdf abstract: In this paper we address the issue of developing an interoperable infrastructure for language resources and technologies. In our approach, called UFRA, we extend the Federate Database Architecture System adding typical functionalities caming from UIMA. In this way, we capitalize the advantages of a federated architecture, such as autonomy, heterogeneity and distribution of components, monitored by a central authority responsible for checking both the integration of components and user rights on performing different tasks. We use the UIMA approach to manage and define one common front-end, enabling users and clients to query, retrieve and use language resources and technologies. The purpose of this paper is to show how UIMA leads from a Federated Database Architecture to a Federated Resource Architecture, adding to a registry of available components both static resources such as lexicons and corpora and dynamic ones such as tools and general purpose language technologies. At the end of the paper, we present a case-study that adopts this framework to integrate the SIMPLE lexicon and TIMEML annotation guidelines to tag natural language texts. |

|  |
| --- |
| Nr 54 |
| Tipologia prodotto Contributo in Conferenza o Convegno (Senza ISBN) |
| Titolo Using LMF to Shape a Lexicon for the Biomedical Domain |
| Descrizione |
| Elenco autori Monachini M.; Quochi V.; Del Gratta R.; Calzolari N. |
| Ruolo svolto |
| anno pubblicazione 2008 |
| abstract: This paper describes the design, implementation and population of the BioLexicon in the framework of BootStrep, an FP6 project. The BioLexicon (BL) is a lexical resource designed for text mining in the bio-domain. It has been conceived to meet both domain requirements and upcoming ISO standards for lexical representation. The data model and data categories are compliant to the ISO Lexical Markup Framework and the Data Category Registry. The BioLexicon integrates features of lexicons and terminologies: term entries (and variants) derived from existing resources are enriched with linguistic features, including sub-categorization and predicate-argument information, extracted from texts. Thus, it is an extendable resource. Furthermore, the lexical entries will be aligned to concepts in the BioOntology, the ontological resource of the project. The BL implementation is an extensible relational database with automatic population procedures. Population relies on a dedicated input data structure allowing to upload terms and their linguistic properties and "pull-and-push" them in the database. The BioLexicon teaches that the state-of-the-art is mature enough to aim at setting up a standard in this domain. Being conformant to lexical standards, the BioLexicon is interoperable and portable to other areas.In abstract: This paper describes the design, implementation and population of the BioLexicon in the framework of BootStrep, an FP6 project. The BioLexicon (BL) is a lexical resource designed for text mining in the bio-domain. It has been conceived to meet both domain requirements and upcoming ISO standards for lexical representation. The data model and data categories are compliant to the ISO Lexical Markup Framework and the Data Category Registry. The BioLexicon integrates features of lexicons and terminologies: term entries (and variants) derived from existing resources are enriched with linguistic features, including sub-categorization and predicate-argument information, extracted from texts. Thus, it is an extendable resource. Furthermore, the lexical entries will be aligned to concepts in the BioOntology, the ontological resource of the project. The BL implementation is an extensible relational database with automatic population procedures. Population relies on a dedicated input data structure allowing to upload terms and their linguistic properties and \"pull-and-push\" them in the database. The BioLexicon teaches that the state-of-the-art is mature enough to aim at setting up a standard in this domain. Being conformant to lexical standards, the BioLexicon is interoperable and portable to other areas. |

|  |
| --- |
| Nr 55 |
| Tipologia prodotto Deliverable di Progetto |
| Titolo Bio-Lexicon DataBase: Architecture, Concepts and Loading Software |
| Descrizione |
| Elenco autori Del Gratta R., Monachini M., Quochi V., Sassolini E., Calzolari N. |
| Ruolo svolto |
| anno pubblicazione 2007 |
| Altre informazioni |

|  |
| --- |
| Nr 56 |
| Tipologia prodotto Deliverable di Progetto |
| Titolo LocalBioLex: A database framework for biolinguistic research on integrated databases |
| Descrizione |
| Elenco autori Del Gratta R., Toral A., Quochi V., Monachini. M. |
| Ruolo svolto |
| anno pubblicazione 2007 |
| Altre informazioni |

|  |
| --- |
| Nr 57 |
| Tipologia prodotto Deliverable di Progetto |
| Titolo TimeML: An Ontological Mapping onto the UIMA Type Systems |
| Descrizione |
| Elenco autori Del Gratta R., Bartolini R., Caselli T., Enea A., Monachini M., Quochi V., Sassolini V. |
| Ruolo svolto |
| anno pubblicazione 2007 |
| Altre informazioni |

|  |
| --- |
| Nr 58 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Quochi V.; Del Gratta R.; Sassolini E.; Monachini M.; Calzolari N. |
| Titolo Toward a Standard Lexical Resource in the Bio Domain |
| Codice identificativo (ISBN) 978-83-7177-413-3 |
| anno pubblicazione 2007 |
| abstract: The present paper describes a large-scale lexical resource for the biology domain designed both for human and for machine use. This lexicon aims at semantic interoperability and extendability, through the adoption of ISO-LMF standard for lexical representation and through a granular and distributed encoding of relevant information. The first part of this contribution focuses on three aspects of the model that are of particular interest to the biology community: the treatment of term variants, the representation on bio events and the alignment with a domain ontology. The second part of the paper describes the physical implementation of the model: a relational database equipped with a set of automatic uploading procedures. Peculiarity of the BioLexicon is that it combines features of both terminologies and lexicons. A set verbs relevant for the domain is also represented with full details on their syntactic and semantic argument structure. abstract: The present paper describes a large-scale lexical resource for the biology domain designed both for human and for machine use. This lexicon aims at semantic interoperability and extendability, through the adoption of ISO-LMF standard for lexical representation and through a granular and distributed encoding of relevant information. The first part of this contribution focuses on three aspects of the model that are of particular interest to the biology community: the treatment of term variants, the representation on bio events and the alignment with a domain ontology. The second part of the paper describes the physical implementation of the model: a relational database equipped with a set of automatic uploading procedures. Peculiarity of the BioLexicon is that it combines features of both terminologies and lexicons. A set verbs relevant for the domain is also represented with full details on their syntactic and semantic argument structure. |

|  |
| --- |
| Nr 59 |
| Tipologia prodotto Deliverable di Progetto |
| Titolo Bio-Lexicon Model and Preliminary ISO Conformant Data Categories |
| Descrizione |
| Elenco autori Quochi V., Monachini M., Calzolari N., Del Gratta R., Sassolini E. |
| Ruolo svolto |
| anno pubblicazione 2006 |
| Altre informazioni |