Progetto UNICittà
RiScopriamo Bologna attraverso l'Alma Mater

[2019.0490]

README

Data Set Title: **Corpus UniCittà**

Data Set Author/s: **Name Surname** (Affiliation), ORCID (if available);

- **Zotti Valeria (Dipartimento di Lingue, Letterature e Culture Moderne - Università di Bologna), 0000-0002-4920-4392**
- **Alaman Pano Ana (Dipartimento di Lingue, Letterature e Culture Moderne - Università di Bologna), 0000-0003-3652-8694**
- **Luporini Antonella (Dipartimento di Lingue, Letterature e Culture Moderne - Università di Bologna), 0000-0003-3312-394X**
- **Perotto Monica (Dipartimento di Lingue, Letterature e Culture Moderne - Università di Bologna), 0000-0001-5510-3363**
- **Turci Monica (Dipartimento di Lingue, Letterature e Culture Moderne - Università di Bologna), 0000-0003-4027-3911**

Data Set Contributor/s: **Name Surname** (Affiliation), ORCID (if available);

- **Caruso Carmelo (Dipartimento di Lingue, Letterature e Culture Moderne - Università di Bologna)**

Data Set Contact Person/s: **Name Surname** (Affiliation), ORCID (if available), email;

- **Zotti Valeria (Dipartimento di Lingue, Letterature e Culture Moderne - Università di Bologna), 0000-0002-4920-4392, <u>valeria.zotti@unibo.it</u>**

Data Set License: this data set is distributed under a **Creative Commons CC by NC ND 4.0**

Publication Year: **2023**

The data set consists of:

- "UNICitta_ExploringBolognaThroughAlmaMaterHeritage_2022.cvs"
- "UNICitta_ExploringBolognaThroughAlmaMaterHeritage_2022_Reademe.txt"

Data set Documentation

Abstract

The UNICittà project, funded by the Carisbo Foundation as part of the Rediscovering The City 2019 call, focuses on the collection and dissemination of a corpus of texts on tangible and intangible heritage belonging to the University of Bologna. This multilingual corpus contains texts written in English, French, Russian and Spanish by foreign travellers of the past which have been collected in order to be made available to organisations and institutions for the creation of tourism promotion resources on the University's heritage. The main goal of the corpus is thus to enhance knowledge and access to such artistic and historical heritage that remains unknown to the wide public and international tourists. This project is part of the activities of the Multilingual Lexicon Research Unit operating at CeSLiC research center (https://site.unibo.it/ceslic/it) in the Department of Modern Languages, Literatures and Cultures.

Fields content description:

- **"TYPE" field:** "TEC" used for technical texts; "DIV" used for popular writings; "LIT" used for literary texts.
- **"LANGUAGE" field**: "French", "English", "Spanish", "Russian".
- **"TEXT" field**: a unit of text about an aspect that is considered of interest with regard to the tangible or intangible heritage of the University of Bologna.
- **"LABELS" field:** each unit of text has been "labelled" with regard to categories that have been identified as highly representative of the tangible or intangible heritage of the University of Bologna. Each unit of text is categorized with reference to one or more representative tags chosen from 8 main categories: Artists ("ART"); Books ("LIB"); Places ("LU"); Objects ("OGG"); Artworks ("OPA"); Historical Period ("PE"); Historical Figures ("PERS"); Theme ("TEM"); Book Titles ("TIL"). The categories and sub-categories - when available - are separated by means of two plus signs "++". Finally, labels are separated with a double colon "::". Categories, sub-categories and labels are in Italian.
    - Example of sub-category: "**PE++Secolo::Diciassettesimo-XVII**" (category: "PE" = Historical Period; sub-category = Secolo (Century); label = "Diciassettesimo-XVII" (17th Century)
    - Example without sub-category: "**TEM++Donna**" (Category = "TEM (theme); label = "Donna" (woman)
- **"BIBLIOGRAPHY" field:** provides reference to the text source, the web link and, when available, to the whole text.

File specifics

To create the .cvs file, data has been converted using Excel (version 2212 Office 365) from .xlsx extension to .csv (UTF-8).

- ***Field delimiter « ; »***