Contribution ID: **65**   Type: **Oral Presentation**

A lexicon for social media-based cultural heritage information in crisis situations: a proposal (Remote Presentation)

Wednesday, 27 March 2024 16:30 (30 minutes)

Social media can play a crucial role in disseminating information about cultural heritage if a proper lexicon is available and able to identify valuable data for the management of crises that are caused by either natural or human-induced disasters. A literature review has been conducted, encompassing existing attempts to define terminology within the cultural heritage domain. A Review of the Role of Social Media in Cultural Heritage Sustainability reveals that there is ongoing interest in the investigation of the role and impact of social media platforms on cultural heritage sustainability and culture preservation. However, the lack of published studies concerning terminological resources for cultural heritage (neither generally, nor in the context of social media discussion) and the absence of a lexicon dedicated to detecting cultural heritage-related tweets on social media during crisis events pushed us to investigate such area of research. For such reason we have undertaken the task of creating our lexicon that provides essential information, comprehends the domain, and facilitates further research in the field. In particular, the lexicon has been defined according to keywords that are commonly used on social media for a specific discussion, and are represented in a list of unigram and bigram terms from natural language processing solutions: e.g., culture or ancient site are keywords for cultural heritage discussion, while vandal or property damage are keywords for vandalism discussion. Furthermore, the defined lexicon can be representative of the domain but also accurately reflect the specific vocabulary commonly utilized within social media platforms, such as Twitter.

Developing a representative lexicon is an essential preliminary step in this study because we have to devise a method for identifying Twitter messages that are related to the field of cultural heritage management in crises. The raw datasets have been collected from January 1 to April 27, 2023, with the Twitter API, in the context of the 4CH project (European Competence Centre for the Conservation of Cultural Heritage) that aims at setting up the methodological, procedural, and organizational framework of a Competence Centre able to seamlessly work with a network of national, regional, and local Cultural Institutions. The collected data, despite being downloaded based on keywords, contain numerous irrelevant tweets and are not suitable for investigation within the context of cultural heritage management in crises. Additionally, the lexicon can enhance the utility of machine learning classification algorithms by serving as a reference point for manual labeling and semi-supervised classification techniques. Consequently, they can be applied to other similar datasets of tweets.

Our dataset is extensive and originates from diverse periods, events, and geographical locations. These distinct locations encompass various nations and institutions, each with its distinct interpretations and definitions of culture and its elements. Questions regarding the nature of culture and what constitutes heritage lack general clear answers on an international scope. In addition, we take into account that the texts collected are in English. This implies that users either come from English-speaking countries or, if they come from other regions, communicate in English due to their connection with an international community or a desire to address global issues using an international language. Given this complexity, we have chosen to create a lexicon that provides the most general framework as possible, relying on the documents of The United Nations Educational, Scientific and Cultural Organization (UNESCO) whose vocabulary is assumed to be close to the one we intend to create for cultural heritage.

Primary authors: Dr SOPYRYAEVA, Anastasiya (INFN CNAF); RONCHIERI, Elisabetta (INFN CNAF); Dr ALKHANSA, Ahmad (INFN CNAF); COSTANTINI, Alessandro (INFN-CNAF); Dr BOMBINI, Alessandro (INFN CNAF)

Presenter: RONCHIERI, Elisabetta (INFN CNAF)

Session Classification: Humanities, Arts & Social Sciences Application

Track Classification: Track 4: Social Sciences, Arts and Humanities (SSAH) Applications