**iPres 2022 Proposal Template Instructions**

This is the DOC/DOT template for Panel, Poster, Workshop, Tutorial and Hack-a-thon proposals for iPres 2022. It is derived from the template used at iPres 2019 and loosely based on the [IEEE template](https://www.ieee.org/conferences/publishing/templates.html). If you prefer writing your paper on a collaborative platform, refer to our Google Docs template which can be found on the [iPres 2022 website](https://ipres2022.org/).

Open the .dotx file to use this template in Word / LibreOffice / OpenOffice.

1. This document uses the **Open Sans** font family.
	1. When you’re on Windows, this font should be included in this template document.
	2. On UNIX systems you have to make sure the system has this font installed. You can [download the TTF here](https://www.fontsquirrel.com/fonts/open-sans).
2. Remove this preamble text and replace the body text with the text of your proposal before submitting.
3. Make sure to use the **Styles Pane** (List: styles in current document) or the **Styles Gallery** to format your document elements (titles, author info, headers, references)
4. The author info is organized in a table.
	1. If your submission has less than three authors, remove the unnecessary columns to space out the authors’ info evenly over the page.
	2. If your submission has more than three authors, duplicate the three rows below the existing one to add extra information
	3. In the case where there are 4 authors, remove one column to end up with 4 x 3 cells).
5. Make sure to ‘Capitalize Each Word’ in your Headings, following the conventions in the template.
6. The final submission should be made in Word – do not export to any other format.
7. Before submitting: Make sure to fill out the Document Properties (*File* > *Properties or File > Info*)

**Automated Topic Modelling in Archives Portal Europe**

|  |  |  |
| --- | --- | --- |
| **Kerstin Arnold** | **Marta Musso** | **Kostantinos Stamatis** |
| *Archives Portal Europe**Germany**kerstin.arnold@archivesportaleurope.net* | *Archives Portal Europe**United Kingdom**marta.musso@archivesportaleurope.net**0000-0002-3728-3548*  | *Archives Portal Europe**Greece**kostas.stamatis@archivesportaleurope.net*  |
|  |  |  |

Archives Portal Europe (APE, www.archivesportaleurope.net) is the portal of European archives, an aggregator that connects on a single research point the catalogues and digitised archival material of all archives in and about Europe. It currently hosts material from more than 30 countries and in 24 languages, and from a variety of archival institutions (such as State and city archives, university and parish archives, private institutions, etc). In order to navigate this extremely heterogeneous material, one of the research tools made available by Archives Portal Europe is by “topics”, curated collections in which each archival institution participating into the APE project can tag its documents according to a specific topic. Because topics are maintained manually by the archivists, and because of the vast amount of archival material ingested in the portal, it is impossible to have a comprehensive body of topics that describe the whole of the APE repository.

In this scenario, automated topic detection can be a fundamental tool to guide archival research and to allow archives to be accessible to potentially world-wide users, in a situation where national and linguistics barriers blur, or are re-defined.

This workshop presents the creation of an AI tool for automated topic detection in the APE corpus, a vast, inhomogeneous, and multi-lingual collection of historical archival catalogues – the first such project to be designed for archival descriptions rather than corpora of specific documents.

The development is based on supervised machine learning, with a combination of human inputs in different languages (collectively-created taxonomies for each topic), and the usage of Wikipedia pages to model the relevant vocabulary and entities. The first iteration of the algorithm was tested on a sample of 9 topics in 5 languages, and the second iteration enlarged the sample to 13 topics and 12 languages, for a total of more than 500,000 descriptive units, and it also introduced Boolean operators and wildcards.

The workshop will explain how the tool was built, and will allow users to test it live, gathering feedback on its usability and possible future implementations outside of the specific corpus of Archives Portal Europe.