KB Koninklijke Bibliotheek
Nationale bibliotheek van Nederland

Submissions KB Researcher-in-Residence program 2018

This document contains all submitted abstracts for the KB Researcher-in-residence program 2018.
Please find the full call for abstracts at: https://www.kb.nl/sites/default/files/docs/call_kb_researcher.pdf .

The representation of paintings in news media: diversity, accessibility and cultural value

News media have always been an important source of information, also to access cultural content. Before the Internet gave access to all the digital images housed in museums and galleries, what paintings appeared in the news? This project will investigate the representation of paintings in news media by analysing their presence in Dutch newspapers and radio news bulletins. Several aspects will be looked at, including the variety and diversity of painters, the context within which the painting was talked about, and the characteristics of the painting (and if an image was present). Results will signal geographic diversity and degree of accessibility to art while reflecting the changing cultural perception and value of paintings. In order to do that, the KB Lab tools will be used, assisted by the KB team, to analyse the KBK-1M Dataset, the searchable newspaper items, and the radio news bulletins.

Back in the Summer of 1869: multimodal and contextual enrichment of musical mentionings in newspapers

Over the past millennia, music has always been performed and listened to by mankind. In the same way as contemporary popular music reflects contemporary sociocultural identities, the genre currently known as Western classical music reflects cultural life in the past centuries. However, because of data sparsity, both caused by smaller audience numbers and inconsistent or incomplete musical item descriptions, we cannot make use of this information in a digitally scalable or systematic way yet. In newspapers, large-scale historical information on daily topics of significance can be found. This also includes mentionings of musical entities. In this project, I want to identify such mentionings, and use them to enable richer, more comprehensive, and more scalable insight into cultural life over the past centuries. Musical entities will be found in two ways: by building forth on existing functionality at the KB for named entity identification and linking, and by developing own matching methods for flexible matching of partial musical item descriptions. The entities will not only be cross-linked with DBpedia, but also with the catalogue of Muziekweb (Centrale Discotheek Rotterdam), containing all recorded music that ever was released in The Netherlands. This way, multimodal and contextual enrichment of both newpaper topics and musical items can be achieved. Thanks to the continuous reinterpretation of classical works, we can acoustically get impressions today of historical periods that were significant before the recording era. In addition, this work will help in shedding light on repertoire that used to be important, but risks being forgotten today

EpIC: the Evolution of the Image of Cities as presented by Dutch local newspapers.

This project aims to build on the geographic dimension present in newspaper articles, as they can be found in Delpher, and uses this geographic information to understand the evolution of relations between places in the Netherlands over time. Reporting about cities has been instrumental in shaping the image of cities, which in turn may have had an effect on the spatial behaviour of people and firms. This research involves three steps/challenges. First, the frequency of reporting about other places by (local) newspapers demonstrates the (perceived) strength of a relation between the place mentioned in the report and the place where the local newspaper is published. This step requires Named-Entity Recognition techniques. This allows to show the evolution of the spatial organisation of the Netherlands through time. Second, we identify what kind of information about other places is being transferred. This involves classifying the articles using Machine Learning, and judging how positive or negative the reporting about other cities is using Sentiment Analysis. Third, the aim is to link this data to spatial behaviour by exploring the hypothesis that cities that are covered more often, and in more positive ways, have been able to attract more inhabitants and firms, whereas the opposite may be true for cities that tended to get a 'bad rap'. This involves statistical modelling, using historic population and employment counts. Efforts will be made to make the geographical tools and datasets derived from this project available to other researchers.

A Rich Knowledge Graph as a Benchmark for Deep Learning on Heterogeneous Data

A key challenge in Artificial Intelligence is integration. If we have a strong algorithm to understand images, and a strong algorithm to understand natural language, how can we combine the two to create a system that can integrate knowledge from both a newspaper image, and the caption accompanying it? What if the types of information available differ from one situation to the next. How to we build a system that uses visual information when it is available, but still functions when it isn't? In [3], it is suggested that to solve this problem, we need deep learning architectures that effectively consume knowledge graphs. To fulfil this promise, we are faced with a Catch-22. To develop such models, we require publicly available, highly curated knowledge graphs containing diverse, rich multimedia information. However, until the envisioned machine learning models exist, no knowledge graphs will be designed with this use case in mind. We propose to use the collection of the Koninklijke Bibliotheek to create the first such knowledge graph: a single database containing a selection of the KB's data, including images, natural language, and audio. For this knowledge graph, we will provide a set of highly challenging machine learning tasks, requiring a deep analysis of the data in different media to solve.

A Rich Knowledge Graph as a Benchmark for Deep Learning on Heterogeneous Data