

DiSCourse Seminar

The Digital Science Center and the Department of Computer Science would like to invite you to the following guest lecture:

Antoine Doucet
University of La Rochelle

Information Extraction from Noisy Text Output

Many documents can only be made accessible for automatic analysis in the form of digitised images. This is particularly the case for any historical or handwritten document, but also for many native digital documents, which have been converted into image form for various reasons (e.g. file conversion or passing through an analog form, e.g., to insert a handwritten signature, send by post, etc.). This lecture will present recent advances in AI and automatic language processing enabling this type of corpus to be analysed in a way that is robust to OCR errors. For example, we will show how, as part of the NewsEye project, we were able to create the state of the art in cross-lingual recognition and disambiguation of named entities (place names, but also names of people and organisations) in old press corpora written in 4 languages between 1850 and 1950, despite particularly degraded corpora. This type of result paves the way for indexing at an advanced semantic level, as well as for large-scale analysis, which can in particular overcome (linguistic) borders.

About the speaker

[Antoine Doucet](#) is a professor at the L3i laboratory at La Rochelle University, where he is head of the "Images and Content" research team. At the intersection of information retrieval, automatic language processing, textual data mining and artificial intelligence, his research focuses on developing methods that can be adapted to very large collections of documents. He'll be visiting the University of Innsbruck in August and September 2024 as part of the LFUI Guest Professorship which is supported by the Circle of Supporters (Förderkreis 1669) and International Services.

Date, Time, Place:

Friday, 13 September 2024, 12:00 (CEST), hybrid

Participants are invited to join the event at the Digital Science Center, Innrain 15, Open Space Area (1st floor) *or* online via [Big Blue Button](#).