



DOCTORATS
INDUSTRIALS



EL PLA DE
DOCTORATS
INDUSTRIALS

PROJECTE DE DOCTORAT INDUSTRIAL EXPEDIENT 2016 DI 087

DADES DE L'EMPRESA I DE L'ENTORN ACADÈMIC

Títol del projecte

Making historical archives searchable. Multiwriter word recognition and spotting using transfer learning models

Empresa

Qidenus Group GmbH

Responsable de l'empresa

Martin Micko Micko

Universitat

Universitat Autònoma de Barcelona

Director/a de tesi

Alicia Fornés Bisquerra

Treballador/a de l'empresa i doctorand/a

Lei Kang

BREU DESCRIPCIÓ DEL PROJECTE DE RECERCA

Historical documents reflect the identities and contexts of the past, and the access to their contents allows citizens at large to know their individual and collective memory. Search centered at people is very important in historical research, including family history and genealogical research. Genealogical documents such as birth, marriage or death certificates or census records contain information that reflect a picture of a historical context: a person's life, an event, a location at some period of time. The CVC and Qidenus are currently working in the development of computational methods for the massive analysis of digitized historical demographic documents.

In order to extract the information contained in demographic manuscript documents, it is necessary to investigate on semantic annotation techniques to associate a semantic meaning to the words of the document. In particular, it is necessary the detection of the so-called named entities and their associated semantic category (concept) such as a person's name, last name, place, date, occupation, etc. Thus, the information contained in the documents can be extracted labeled and stored in databases, and hence making the information accessible and searchable.

Methodologically, the thesis will focus on techniques that learn a joint embedding space that associates visual features extracted from images to concepts. We plan to use deep neural networks to associate semantic categories to named entities.



Generalitat de Catalunya
Departament d'Empresa i Coneixement
Secretaria d'Universitats i Recerca



Agència
de Gestió
d'Ajuts
Universitaris
i de Recerca