CRIM IS PART OF A VAST TECHNOLOGICAL PROJECT ON INDIGENOUS LANGUAGES IN CANADA!

Montréal, December 5, 2018 – CRIM is proud to announce the beginning of a long-term collaboration with the National Research Council of Canada (NRC) on a project to support the revitalization and preservation of Indigenous languages through text and speech-based technologies. CRIM will use its expertise to tailor its speech recognition technologies to Indigenous languages.

The Canadian Indigenous languages technology project is an NRC initiative funded by the federal government in Budget 2017. CRIM will also work with researchers from Carleton University and the University of Alberta who have been involved with Indigenous communities for several years. The teams will work in collaboration with Indigenous community-based organizations and Indigenous communities across Canada.

CRIM's role

- CRIM will carry out **two main projects** that will serve as a basis for the development of a dozen systems related to speech recognition and adapted to the target languages.

- The first project will lead to the development of **segmentation tools for audio recordings**, making it possible to distinguish speech from music or noise, to identify the languages spoken and to separate the different speakers. These tools will facilitate the annotation and transcription of content to accelerate the documentation of existing corpora for each language.

- The second project focuses on the design of an **indexing tool** to identify and organize existing audio content for each of the target languages. This tool will make it easy to navigate through the many recordings available, for instance to find video footage that discusses a specific topic or to discover how certain expressions are used.

A crucial project, both for its social impact and for scientific progress

Researcher Roland Kuhn, head of the Canadian Indigenous languages technology project at the NRC, describes why this joint research with CRIM is important for Canada's Indigenous communities:

"There are thousands of hours of recordings in Indigenous languages, but unfortunately these are rarely annotated or indexed due to a lack of appropriate technology to do so: this unannotated corpus is constantly growing. This is frustrating for members of the affected communities, many of whom would like to be able to use keywords to search for records that are relevant to their current needs. A person should not have to listen to 10,000 hours of audio in order to find a recording about a traditional ceremony in their community, for example. Our team is very pleased to collaborate with CRIM, an organization that is internationally recognized for its text and speech-based technologies, on this important cultural and linguistic challenge."

CRIM part of a vast technological project
on Indigenous languages in Canada
December 5, 2018

The structure of Indigenous languages is very different from that of English or French, which makes many of CRIM's current speech recognition methods insufficient. This poses a major challenge and requires the development of new approaches that will certainly lead to innovations in the field.

Both parties aim to provide high-quality technological tools for the teaching, valorization and preservation of languages for Indigenous communities. In addition, CRIM experts hope that the methods developed for the languages targeted at this stage of the project (Inuktitut and Cree) will be applicable to several more of the 70 Indigenous languages spoken in Canada.

About CRIM

CRIM is an applied research and expertise centre in information technology, dedicated to making organizations more effective and competitive through the development of innovative technology and the transfer of leading edge know-how, while contributing to scientific advancement.

It helps organizations, primarily SMBs, demystify and gain access to leading-edge technology, such as artificial intelligence, to efficiently address the technological challenges they face. Its IT researchers and professionals develop a wide array of applications in diverse areas and work in such fields of expertise as machine learning, computer vision, speech recognition, automatic natural language processing, data science and operational research.

CRIM is a non-profit organization whose neutrality and strong network make it an indispensable resource. Its work is in line with the policies and strategies of its major financial partner, the ministère de l'Économie et de l'Innovation.

– 30 –

Link

Project description and list of partners (NRC website)

Contact and information

Antonia Leney-Granger, communications agent
CRIM – Computer Research Institute of Montréal
medias@crim.ca | www.crim.ca | 514 840-7493