Assisting Digital Sensitivity Review of Governments Records

Freedom of Information (FOI) laws legislate for the principle that government documents should be opened to the public for the general good. However, many government documents contain sensitive information that is exempt from release under these laws where the damage from release will exceed the benefit. Therefore, government documents must be sensitivity reviewed prior to release, to identify, and handle such information appropriately. With the adoption of born-digital documents, such as email, traditional (paper-based) sensitivity review processes are no longer suitable and there is a need for automatic sensitivity classification to assist digital sensitivity review. In the first half this talk, Tim Gollins from the National Records of Scotland will provide an overview of the challenges of sensitivity review of government records, from both UK and Scottish legal perspectives. Having defined these challenges, in the second half of this talk, Craig Macdonald from the University of Glasgow will describe work on assistive classifiers that can make predictions about the sensitivity of records, and thereby assist reviewers in assessing sensitivity. He will provide an overview of recent work in incorporating textual classifiers, as well as classifiers that can use evidence about syntactic similarities based on Parts-of-Speech, and semantic similarities obtained from word embeddings. He will also describe classifiers that can learn in an active manner as reviewers assess records.