

Thesis advisor: Dr. Séamus Lawless

Neil Hickey

## *Evaluation of Test Collections Generated from Transaction Logs in Digital Humanities*

### ABSTRACT

Understanding relevance in Information Retrieval is a difficult task. Measuring the effectiveness of an information retrieval system is predicated on knowing whether a document is relevant to a search query, or not. However, relevance is subjective, personal and dynamic. Digital Humanities poses particular challenges with regards to relevance, due to smaller sized collections and the nature of these collections, amongst other factors.

This dissertation proposes and evaluates a pipeline which hopes to generate a test collection, which is used to evaluate these information retrieval systems, from transaction logs for digital libraries in the domain of Digital Humanities.

The proposed experiment is implemented using methodologies from *transaction log analysis*, and the generated test collection is evaluated through a comparison of this generated collection and expected results, as seen in literature, for two real-world digital library datasets.

This dissertation concludes that it was not possible to derive a test collection by using binary judgments of relevance from transaction logs, for the particular collections examined.