Determining and Visualising E-mail Subsets to Support E-discovery

Mithileysz Sathiyanarayanan, Cagatay Turkay
giCentre, City, University of London

Introduction

Electronic discovery (E-discovery): Investigating emails collected over a period of time, manually, is a strenuous process and the current keyword search is very expensive.

So, there is a great need to determine, visualise and understand whether email subsets are normal or abnormal, pertinent or privileged, relevant (interesting) or immaterial in a quick time.

We proposed a multi-modal and multi-level approach which will generate automated visual representations using a manual search facility.

Design Approach

Level 1: Overview
Represents "anomaly" and "relevance" in data
Features:
- Choose time, individuals and keywords independently
- Data selection and filtering

Level 2: Details
Represents "interestness" in data
Features:
- Choose time-time, individuals-individuals and keywords-keywords combination
- Data selection and filtering
- User interaction

Level 3: Particulars
Represents "Priviledgeness" in data
Features:
- Choose time-individuals, time-keywords and individuals-keywords combination
- Data selection and filtering
- User interaction

ONGOING WORK

- The prototype will be turned into a simple, powerful and analyst-friendly visualisation tool that will be tangible and feasible to use in E-discovery investigations.
- Text analytics such as automated Named Entity Recognition or Classification of Email categories will be considered in order to provide valuable data preprocessing/analysis.
- Text visualization will be considered in order to provide effective views for the processed data.
- The complete version of the tool will have user testing using Amazon Turk to evaluate the visualisation choices.

Discover insight from your data and automate workflows.
Redsift now in Public beta. Use QR Code for more details.