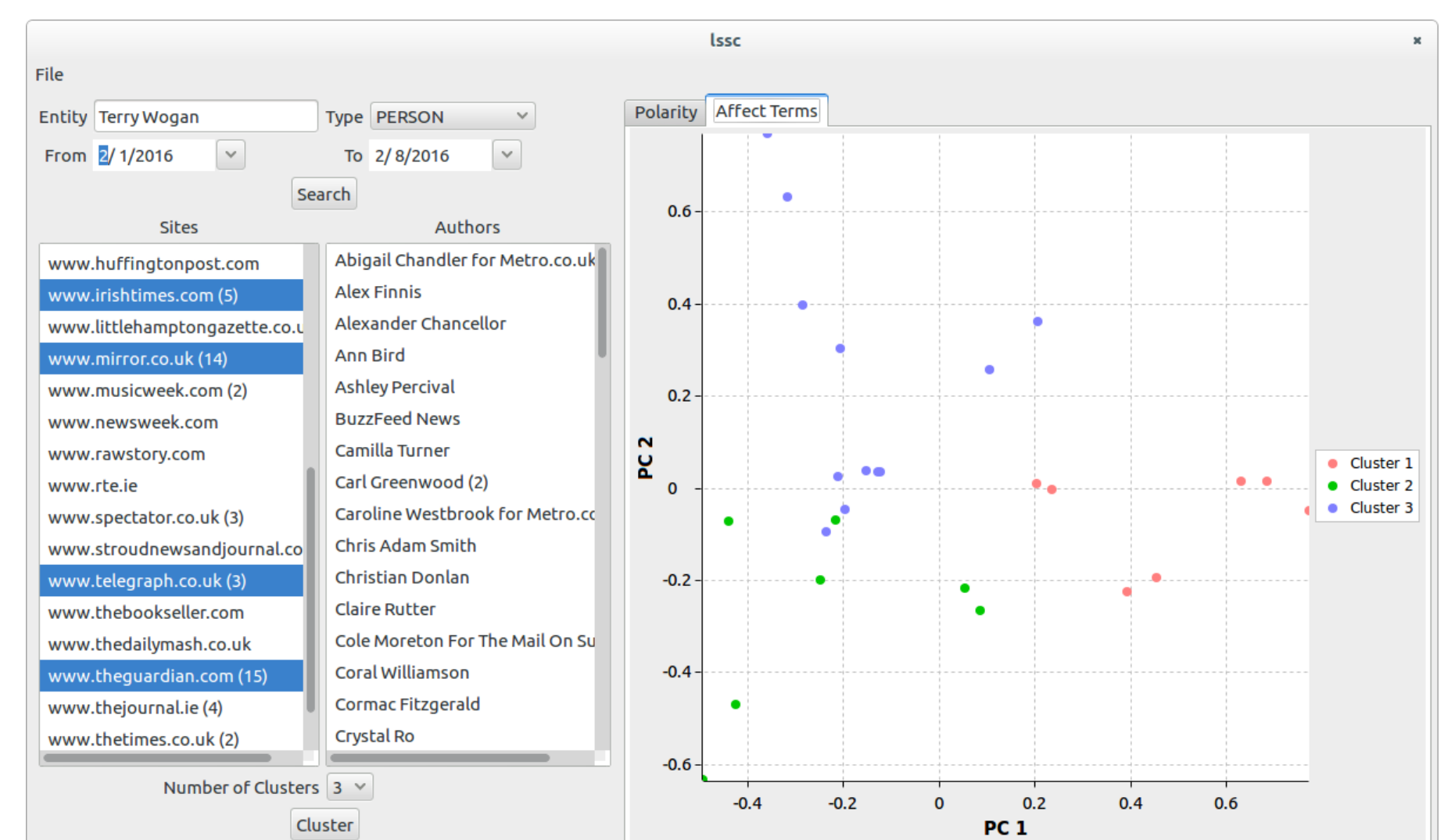## MARKET NEED

There has been enormous growth in the field of sentiment analysis in recent years. CeADAR recently delivered Next-Generation Sentiment, software which provides a sophisticated measure of sentiment and a better understanding of the emotional composition of sentiment. While that solution allows the user to develop a very nuanced understanding of the sentiment expressed in a document, it lacks a means to compare the results of sentiment analysis from one document to another. We now introduce the complementary software Large-Scale Sentiment Clustering, which allows the clustering of documents based on sentiment and the identification of exceptions to the mainstream opinion in a set of documents.
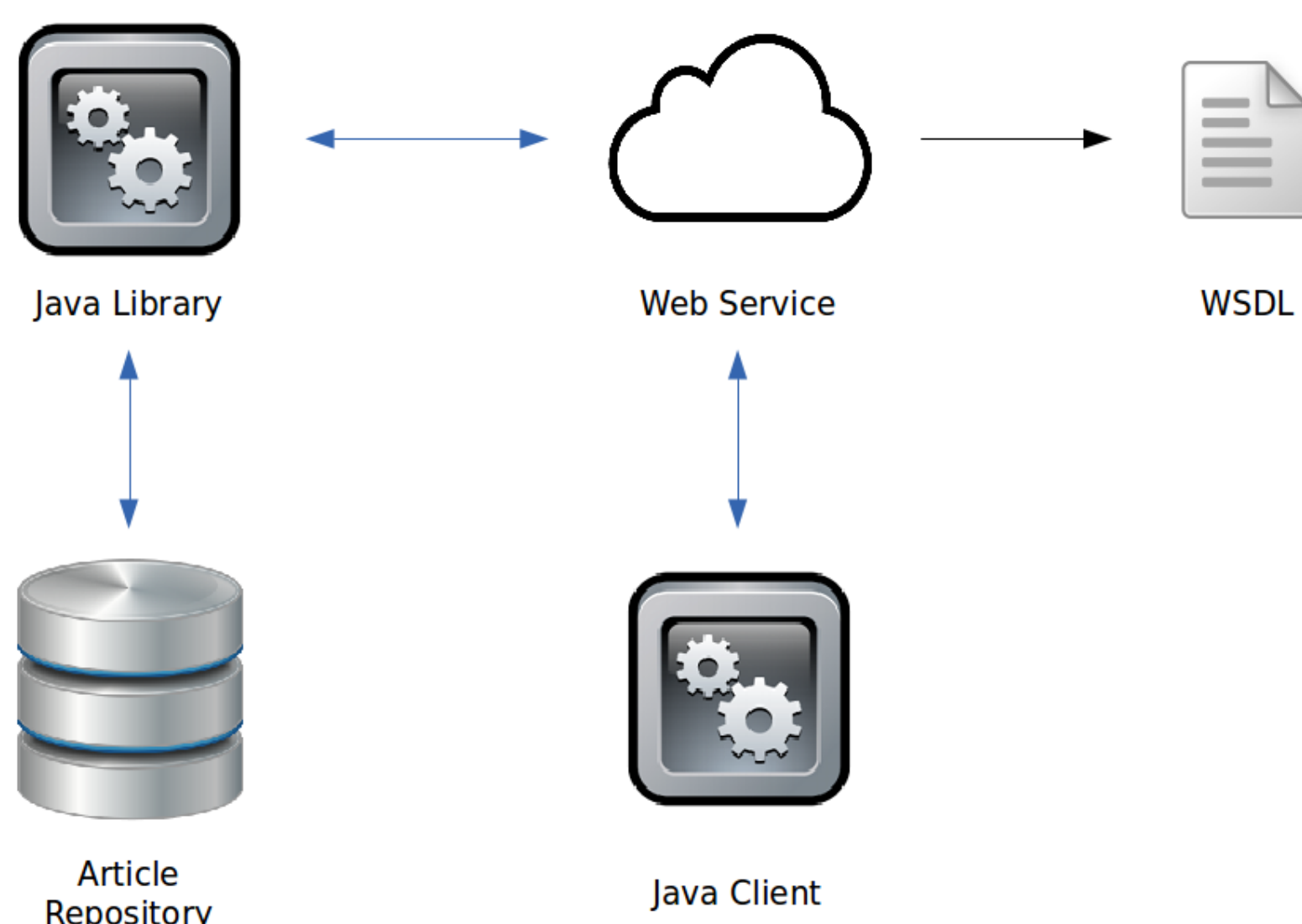
TECHNOLOGY SOLUTION

We have developed a software library which operates on a repository of several million digitised newspaper and magazine articles (provided by our partner Scredible). The software combines advanced sentiment analysis and clustering techniques in a manner which allows the processing of large numbers of articles in parallel. The approach can be applied to each article as a whole or can be targeted at particular entities of interest which feature in the articles.

APPLICABILITY

The Large-Scale Sentiment Clustering software is suitable for any organisation which requires both a detailed understanding of the sentiment expressed in documents of interest to them and an understanding of how that sentiment is distributed across the set of documents. The document repository can be pre-populated by the user, or built up incrementally through continued use of the software, or both. The software library can be incorporated in Java programs directly or, by using the web service included, in programs written in other programming languages.

Clustering results

RESEARCH TEAM

The research team at UCC consists of Professor Barry O'Sullivan and Dr Liam O'Toole.

System architecture

TECHNOLOGY CENTRE
ENTERPRISE IRELAND
IDA IRELAND SUPPORTED