Design Concept

4.1 Problem Definition

We are presented with the ISEBEL project, a digital archive of stories from belief legends found in three well known digital collections by Evald Tang Kristensen from Denmark (etkspace), Richard Wossidlo from Mechlenburg (wossidia) and several collectors and narrators from the Netherlands (verhaalenbank). These databases are made up of stories originating from different source. Stories are composed by different authors and spread across many papers in the database, stories in these papers have facts and contents which are related and also an author may contribute related ideas to content of different papers. Thus, these databases are stock with varieties of stories, with some stories so inter-related such that comprehensive information cannot be found in a single document, stories have to be compiled from documents all across the archive. Facts and contents of the stories may be related either by the stories itself or by author and co-authorship.

Therefore, we are faced with the challenge of modeling and implementing a framework to present the users of ISEBEL search system with a more relevant search result that effectively gives results of stories, related stories, possible stories that will be interconnected in the feature and a way to visualize stories inter-relatedness through the use of data mining techniques.

4.1 Data Extraction process

One of the key steps towards a successful data mining is the availability of data. In this research, the data the XML story data to be used is harvested from the ISEBEL project using OAI-PMH.