Statistics  
   
Statistics  
Statistics (from German: Statistik, orig. "description of a state, a country") is the discipline that concerns  
the collection, organization, analysis, interpretation, and presentation of data. In applying statistics to a  
scientific, industrial, or social problem, it is conventional to begin with a statistical population or a  
statistical model to be studied. Populations can be diverse groups of people or objects such as "all  
people living in a country" or "every atom compo

## Wikipedia Excerpt: Document classification

Document classification or document categorization is a problem in library science, information science and computer science. The task is to assign a document to one or more classes or categories. This may be done "manually" (or "intellectually") or algorithmically. The intellectual classification of documents has mostly been the province of library science, while the algorithmic classification of documents is mainly in information science and computer science. The problems are overlapping, however, and there is therefore interdisciplinary research on document classification.  
The documents to be classified may be texts, images, music, etc. Each kind of document possesses its special classification problems. When not otherwise specified, text classification is implied.  
Documents may be classified according to their subjects or according to other attributes (such as document type, author, printing year etc.). In the rest of this article only subject classification is considered. There ar