Data warehousing techniques are often used as one of the principal components of Decision Support Systems, and decision support systems are computer-based information systems that businesses use to make decisions. They provide the tools necessary to perform analysis and arrive at a decision.

DWs are not only used in businesses ..

The project I worked on for this course consisted in building a data warehouse of places mentioned books. The books are collected from the Gutenberg Canada website which provides them free of charge. The books are often, books that have had their copyright expired 50 years after the author’s death. The Gutenberg Canada Website provides books in Text Files and other formats.

This project is inspired from the LitOLAP project.

LitOLAP is a project that seeks to apply business intelligence techniques and data warehousing in the domain of text processing. And, the data warehouse I am building is also part of this project. LitOLAP seeks to allow answering questions regarding literary text that can allow a research answer questions over an author’s style, or particularities about a book among other things. Examples of queries that maybe presented are: what the most common words are, or common pairs or n-grams which are a group words that appear in a particular order, or co-occurrences words which are words that are found together without a particular order. The goal of LitOLAP then is to facilitate the analysis of texts to a domain expert.

To do this, LitOLAP builds different OLAP cubes, by breaking up the text in certain blocks or tokens, and performs a measure.

An important part of designing a Data Warehouse is defining in the multi-dimensional model, which determines how the data is structured in your data warehouse.

The MD model divides information between measures and context. Measures are usually numerical data that can be summarized and are stored as facts. The context is represented by the dimensions and can help us answer the questions of who, what when, where, why and how of a measurement.

Measures can be additive, semi-additive and non-additive.

We often hear of the Entity Relation model, which is also used to define how the data is structured in a relational database.

Author by Id  
When we have the author as a separate dimension we can better understand manage the many to many relationship between authors and books.  However because most books are not written by the same author the cube would become even sparser.  
  
Another important thing to mention is that the etl process is done in two passes one in which URLs are extracted and downloaded and

Data warehousing techniques are often used as one of the principal components of Decision Support Systems, and decision support systems are computer-based information systems that businesses use to make decisions. They provide