

# Chapter 56

## Semantic Interation, Text Mining, Tools and Technologies

**Chandrakant Ekkirala**  
*Cognizant Technologies Limited, India*

### ABSTRACT

*Semantic technologies have gained prominence over the last several years. Semantic technologies are explored in detail and semantic integration of data will be outlined. The various data integration techniques and approaches will also be touched upon. Text Mining, different associated algorithms and the various tools and technologies used in text mining will be enumerated in detail. The chapter will have the following sections – 1. Data Integration Techniques • Data Integration Technique – Extraction, Transformation and Loading (ETL) • Data Integration Technique – Data Federation 2. Data Integration Approaches • Need Based Data Integration • Periodic Data Integration • Continuous Data Integration 3. Semantic Integration 4. Semantic Technologies 5. Semantic Web Technologies 6. Text Mining 7. Text Mining Algorithms 8. Tools and Technologies for Text Mining*

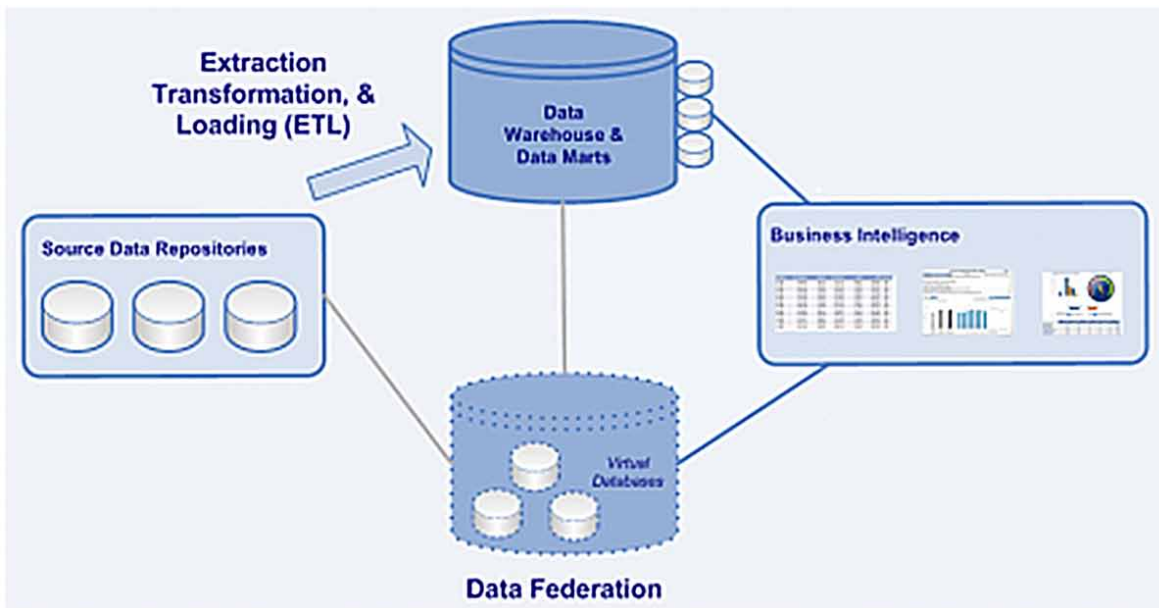
### INTRODUCTION

#### Data Integration Techniques

Data integration is a fundamental, yet deceptively challenging, component of any organization's business intelligence and data warehousing strategy. Data integration involves combining data residing in different data repositories and providing business users with a unified view of this data. In addition, companies face a challenge of ensuring that data being reported is current and up-to-date. Companies are now increasingly incorporating both traditional batch-oriented techniques for query performance and real-time data integration to eliminate the annoyance of out-of-date data. The top batch-oriented technique that companies utilize is known as ETL while one of the popular real-time techniques is known as Data Federation.

DOI: 10.4018/978-1-5225-1759-7.ch056

Figure 1. Data Integration Techniques – ETL and Data Federation



### Data Integration Technique: Extraction, Transformation and Loading (ETL)

The term ETL which stands for extraction, transformation, & loading is a batch or scheduled data integration processes that includes extracting data from their operational or external data sources, transforming the data into an appropriate format, and loading the data into a data warehouse repository. ETL enables physical movement of data from source to target data repository. The first step, extraction, is to collect or grab data from its source(s). The second step, transformation, is to convert, reformat, cleanse data into format that can be used by the target database. Finally the last step, loading, is import the transformed data into a target database, data warehouse, or a data mart. A data warehouse holds very detailed information with multiple subject areas and works towards integrating all the data sources. A data mart usually holds more summarized data and often holds only one subject area.

#### ETL Step 1: Extraction

The extraction step of an ETL process involves connecting to the source systems, and both selecting and collecting the necessary data needed for analytical processing within the data warehouse or data mart. Usually data is consolidated from numerous, disparate source systems that may store the data in a different format. Thus the extraction process must convert the data into a format suitable for transformation processing. The complexity of the extraction process may vary and it depends on the type and amount of source data.

16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/semantic-interation-text-mining-tools-and-technologies/173385](http://www.igi-global.com/chapter/semantic-interation-text-mining-tools-and-technologies/173385)

## Related Content

---

### The Importance of Interface Agent Characteristics from the End-User Perspective

Alexander Serenko (2008). *Intelligent Information Technologies and Applications* (pp. 137-151).

[www.irma-international.org/chapter/importance-interface-agent-characteristics-end/24263](http://www.irma-international.org/chapter/importance-interface-agent-characteristics-end/24263)

### Detection and Classification of Leukocytes in Blood Smear Images: State of the Art and Challenges

Renuka Veerappa Tali, Surekha Borraand Mufti Mahmud (2021). *International Journal of Ambient Computing and Intelligence* (pp. 111-139).

[www.irma-international.org/article/detection-and-classification-of-leukocytes-in-blood-smear-images/275761](http://www.irma-international.org/article/detection-and-classification-of-leukocytes-in-blood-smear-images/275761)

### Adoption of Machine Learning With Adaptive Approach for Securing CPS

Rama Mercy Sam Sigamani (2020). *Handbook of Research on Machine and Deep Learning Applications for Cyber Security* (pp. 388-415).

[www.irma-international.org/chapter/adoption-of-machine-learning-with-adaptive-approach-for-securing-cps/235051](http://www.irma-international.org/chapter/adoption-of-machine-learning-with-adaptive-approach-for-securing-cps/235051)

### Designing and Modelling of Delta Wing Genetic-Based Prediction Model

Arun M. P., Satheesh M. and J. Edwin Raja Dhas (2021). *International Journal of Ambient Computing and Intelligence* (pp. 159-183).

[www.irma-international.org/article/designing-and-modelling-of-delta-wing-genetic-based-prediction-model/272043](http://www.irma-international.org/article/designing-and-modelling-of-delta-wing-genetic-based-prediction-model/272043)

### Multifaceted Applications of Data Mining, Business Intelligence, and Knowledge Management

Kijpokin Kasemsap (2018). *Intelligent Systems: Concepts, Methodologies, Tools, and Applications* (pp. 810-825).

[www.irma-international.org/chapter/multifaceted-applications-of-data-mining-business-intelligence-and-knowledge-management/205810](http://www.irma-international.org/chapter/multifaceted-applications-of-data-mining-business-intelligence-and-knowledge-management/205810)