# We are IntechOpen,
# the world's leading publisher of
# Open Access books
# Built by scientists, for scientists

## 6,900
Open access books available

## 186,000
International authors and editors

## 200M
Downloads

Our authors are among the

## 154
Countries delivered to

## TOP 1%
most cited scientists

## 12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

**WEB OF SCIENCE**™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

# Interested in publishing with us?
# Contact book.department@intechopen.com

# Information Gathering and Classification for Collaborative Logistics Decision Making

José Ceroni and Rodrigo Alfaro
*Pontificia Universidad Católica de Valparaíso*
*Chile*

## 1. Introduction

Collaboration is perceived as a powerful tool for companies to deal with an increasingly demanding global economic environment. Collaboration impact upon the logistics processes is analyzed in this chapter. Logistics naturally implies the concurrency of a set of companies being part of products and services value chain. However, collaborative logistics imposes new challenges that effectively faced by companies and government institutions will produce the expected return. Such challenges relate to effective decision making process involving logistics activities planning, scheduling, control, and coordination at different companies being part of the logistics network. These decision making processes require the effective capability to capture, process, and analyze information on processes, partners, environment, regulations, etc. That information is, of course, dynamic, and it is available in a scattered way all through a variety of sources, periodicity, and formats. Effective and efficient tools for capturing, classifying, processing, and report useful information for supporting the collaborative logistics decision making processes are required. Large amounts of digital text available on the web contain useful information for enabling collaborative logistics. The amount of digital text it is expected to increase significantly in the near future, making the development of data analysis applications an urgent need. Information gathering has been traditionally faced integrating systems and databases at the various institutions and companies participating in the logistics network. This approach has at least two difficulties that have not been solved properly: 1) It requires the data be available in structured databases and stored in terms of the same attributes, and 2) Organizations need to provide access to their systems for obtaining and processing the data. An alternative approach is using data available in the web as input to automatic text classifiers implemented with machine learning techniques. Automatic text classification (or categorization) is defined as the assignment of a Boolean value to each pair $(d_j, c_i)$ belonging to the set $D \times C$, where $D$ is the domain of documents and $C = \{c_1, \ldots c_{|C|}\}$ is the set of predefined labels. Binary classification is the most simple and widely studied case, in which a document is classified into one of two mutually exclusive categories or classes. Document representation has a high impact on the task of classification. Some elements used for representing documents are: N-grams, single-word, phrases, or logical terms and statements. The vector space model is one of the most widely used models for ad-hoc information retrieval, mainly because of its conceptual simplicity and the appeal of the

underlying metaphor of using spatial proximity for semantic proximity. In the vector space model (VSM), the contents of a document are represented by a vector in the term space: $d = \{w_1;...,w_k\}$, where $k$ is the size of the term (feature) set. We propose a solution based on automatic text classifiers using web documents (such as twitters or microblogs) for supporting collaborative logistics decisions making processes such as planning, identifying, and creating value, enabling the flow of value stream, allowing customers to pull value uninterrupted, responding to unpredictable change, or forming tactical and virtual partnerships. Testing of the proposed solution was carried out with a prototype implementing the automatic text classifier for the above decision making processes.

## 2. Modeling the supply chain

For the last decade a new generation of reference frameworks has been developed in response to developments in supply-chain management. Such frameworks have been devised to model inter-company relationships as a basis for inter-enterprise software development with examples such as the Supply Chain Operations Reference-model (SCOR) and the Collaborative Planning, Forecasting, and Replenishment initiative (CPFR). Industry applications in Customer Relationships Management (CRM), Advanced Planning Systems (APS), Product Life Cycle Management (PLM), and other enterprise areas have been further developed to exploit and foster these frameworks (Hvolby & Trienekens, 2010).

### 2.1 Business systems integration frameworks
Next, four of the major supply chains modeling frameworks are presented briefly:
- Collaborative Planning, Forecasting, and Replenishment (CPFR)
- ISA-S95 standards for enterprise and manufacturing integration
- Integration Specifications developed by Open Applications Group (OAG)
- Supply Chain Operations Reference-model (SCOR)

CPFR has developed a set of reference business processes, which can be used for collaboration on a number of buyer/seller functions towards overall efficiency in the supply chain. The framework focuses on efficient and effective retailer–manufacturer relationships to support high consumer value. The key supply-chain functions in CPFR are planning, forecasting and replenishment. On manufacturer and retailer level specific business processes are defined (such as logistics and execution monitoring for the manufacturer and store execution and supplier scorecard for the retailer). On the interface level processes such as joint business plan, sales forecasting and order-planning are defined and worked out. Their XML specifications have been integrated with the broader set of EAN-UCC XML specifications endorsed by the Global Commerce Initiative (GCI) to ensure full coverage of the CPFR process without creating overlapping or redundant message formats. The existing core EAN-UCC messages for item synchronisation, party (trading partner) synchronisation, purchase order, invoice, dispatch (shipment notice) and other information have been augmented with the CPFR product activity, forecast and other transactions.

ISA-S95 (IEC62264) addresses the interface or exchange of data within the enterprise systems (planning, scheduling and procurement) and the production management systems (production dispatching and execution). The ISA-S95 standards are based upon 4 functional levels: business planning and logistics (level 4), manufacturing operations management (level 3), manufacturing process systems for batch, continuous and discrete control (level 2)

and finally level 1, which e.g. is sensing the production process. The scope of the standard is limited to describing the relevant functions in the enterprise and the control domain and which objects normally exchanged between these domains. The standard consists of three parts: Models and terminology, Object model attributes and Models of Manufacturing Operations. The development is based on the work by Williams (1992) on the Purdue Reference Model (PRM) for Computer Integrated Manufacturing (CIM), but two other works have also a great deal of influence, which are the ISA Sp-88 ''Batch Control' committee and the Mesa International MES context model (Hvolby & Trienekens, 2010).

OAG includes a broad set of XML schemas for sharing business information. OAGIS XML is the most mature XML language today, based on over 10 year's development. It addresses the needs for traditional ERP integration as well as supply-chain management and e-commerce. This specification provides the structure of business documents and additional meta-data, which is required as a part of the application processing. ISA-S95 standardizes a variety of models and terminology that is limited to describing the relevant functions, and serves only as a foundation for implementation elements. OAGI, on the other side, defines implementations elements, but without explicitly defining standardized models and terminology, as ISA-S95. OAGI and ISA-S95 are collaborating on development of integration standards for process, discrete, and mixed-mode manufacturers. ISA-SP95 has started this effort by including a portion of the OAGIS standard in its ANSI/ISA-S95 Part 5 – Business to Manufacturing Transactions standard. Also delivery of the ISAS95 Part 4 – Object Models and Attributes of Manufacturing Operations Management standard has greatly accelerated by leveraging the work of OAGI (Hvolby & Trienekens, 2010).

In 1996 a group of 70 companies founded the Supply-Chain Council (SCC). The aim of the council was to create a Supply Chain Operations Reference-model (SCOR-model) that is branch independent and allows the exchange of information between companies in a supply chain. Furthermore, the SCOR-model was designed to enable companies to compare and learn from companies within and outside their own field. The SCOR-model approach gives these companies the chance to standardize the description of supply chains, which is very useful in order to form a unified understanding of operations and to compare different supply chains (Röder & Tibken, 2006).

The SCOR-model has been developed to describe the business activities associated with all phases of satisfying a customer's demand. The model itself contains several sections and is organized around the five primary management processes of Plan, Source, Make, Deliver and Return (Supply Chain Council, 2010; Corsten, 2001; Schönsleben, 2000). These five management processes are represented in Figure 1.

By describing supply chains using these process building blocks, the model can be used to describe simple supply chains as well as very complex enterprise networks using a common set of definitions. As a result, disparate industries can be linked to describe the depth and breadth of virtually any supply chain. The model is able to describe and provide successfully a basis for process improvement for global projects as well as site- specific projects.

The SCOR-model spans all customer interactions (order entry through paid invoice), all physical material transactions (supplier's supplier to customer's customer, including equipment, supplies, spare parts, bulk product, software, etc.) and all market interactions (from the understanding of aggregate demand to the fulfillment of each order). It does not attempt to describe every business process or activity. Specifically, the model does not address sales and marketing (demand generation), product development, research and development and some elements of post-delivery customer support.

The structure of the SCOR-model includes four levels that represent the path a company takes to improve its supply chain. The four levels are described in Figure 2.
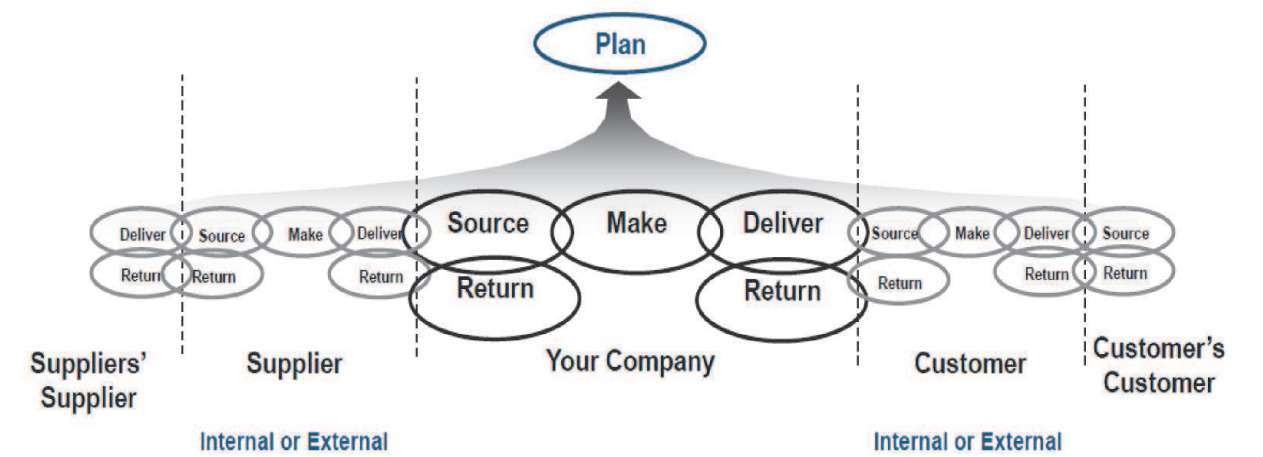


Fig. 1. The five major management processes of SCOR-model (Supply Chain Council, 2010)



| | Level | Comments | Schematic |
|---|---|---|---|
| **Supply Chain Operations Reference-model** | **(1) Top Level (Process Types)** | Level 1 defines the scope and content for the Supply Chain Operations Reference-model. Here the basis of competition performance targets is set. | |
| | **(2) Configuration Level (Process Categories)** | A company's supply chain can be "configured-to-order" at level 2 from 26 core "process categories". Companies implement their operations strategy through the configuration they choose for their supply chain. | |
| | **(3) Process Element Level (Decompose Processes)** | Level 3 defines a company's ability to compete successfully in its chosen markets. It consists of:<br>• Process element definitions<br>• Process element information inputs and outputs<br>• Process performance metrics<br>• Best practices, where applicable<br>• System capabilities required to support best practices<br>• Systems/tools<br>Companies "fine tune" their operations strategy at level 3. | |
| **Not in Scope** | **(4) Implementation Level** | Companies implement specific supply chain management practices at this level. Level 4 defines practices to achieve competitive advantage and to adapt to changing business conditions. | |

Fig. 2. Overview of Supply Chain Operations Reference-model (Röder & Tibken, 2006).

- Level 1. This level provides the definition of the plan, source, make and deliver process types. At this level the company defines its supply chain objectives.
- Level 2. Level 2defines 26 core process categories that are possible components of a supply chain. Organizations can configure their ideal or actual operations by using one or several of the core process categories.
- Level 3. Level 3 provides the information required for successfully planning and setting goals for supply chain improvements. This includes defining process elements, setting target benchmarks, defining best practices and system software capabilities to enable best practices.
- Level 4. This level focuses on the implementation, i.e., putting specific supply chain improvements into action. These are not defined within the industry standard model since implementation can be unique to each company.

## 2.2 Modeling intra- and inter-company supply chains

In order to validate concepts of cooperation and collaboration in inter- and intra-company supply chains and to optimize business processes in existing supply chain structures realistic modeling of supply chains is necessary. The realistic description includes the evaluation of different supply chain structures and configurations as well as different sets of parameters describing production and inventory processes. Therefore, the modeling methodology has to fulfill the following requirements of the industry:

- Requirements concerning the structures of supply chains
  - Description of intra- and inter-company supply chains.
  - Modeling of divergent sourcing and supply structures.
  - Describing company sectors by using logistic processes as source, make, and deliver processes.
  - Modeling different types of inventories (e.g. inventory for incoming goods, inventory for outgoing goods).
- Requirements concerning the material and information flows of supply chains
  - Integrated description of material flows and information flows as well.
  - Modeling of time interdependencies and time lags in material and information flows.
  - Modeling of different types of capacity constraints.

## 2.3 Configuration of intra- and inter-company supply chains

In order to describe the logistic processes and to assess the benefits of an integrated product and process documentation, logistic processes of intra- and inter-company supply chains have to be modeled in detail. Modeling the structures in order to configure supply chains represents the first part of the modeling methodology. The modeling is based on the SCOR-model, which flexibility and especially modularity are very useful for the configuration of structures. SCOR allows describing whole production processes as well as detailed manufacturing processes. The different methods to initiate orders, e.g. make-to-stock, make-to-order, engineer-to-order, have been incorporated as well (Supply Chain Council, 2010; Corsten, 2001). Depending on the focus of modeling, the SCOR-model approach allows a wide range of modeling material and information flows within inter- and intra-company process structures as well as high degrees of adaptability and flexibility to fast changing interdependent processes and structures.

## 3. Considering transportation capacity in the logistics modeling framework

By 2050, the total U.S. population is projected to reach 420 million, a 50 percent increase over 50 years. This growing society will demand higher levels of goods and services, and will rely on the transportation system to access them. In turn, this will cause travel to grow at an even greater rate than the population. As part of an increasingly integrated global economy, the U.S. will see greater pressures on its international gateways and domestic freight distribution network to deliver products and materials to where they are needed. The Nation is faced with a massive increase in passenger and freight travel (NSTPRC, 2008). This particular situation in the United States is assumed to hold for most of the countries with open economies, which depend heavily on exchange of products. As the Revenue Study Commission points out, the cost and consequences of inaction are enormous and lead to:

- Country's transportation system assets will further deteriorate
- Automobile casualties will increase, adding to the 3.3 million lives lost to traffic crashes in the last 100 years
- Congestion will continue to affect every mode of surface transportation for ever-lengthening periods each day, as a result of the mismatch between demand and supply of limited capacity
- Underinvestment in all modes will continue
- America's economic leadership in the world will be jeopardized when we cannot reliably and efficiently move our goods
- Excessive delays in making investments will continue to waste public and private funds
- Transportation policies will remain in conflict with other national policy goals
- Transportation financing will continue to be politicized

To avoid such a disastrous scenario, a capacity evaluation framework was proposed. We propose the framework can be adapted from the U.S. reality to other countries as well. The transportation decision making process is made of many individual steps. Most of these steps are work activities that take place in the technical decision making process. Key decisions are those places in the process where the general work activities need review and approval from higher levels of authority or where consensus needs to be reached among diverse decision makers before the project can advance further. For this reason key decisions most often occur in the policy decision-making process. Key decision points, therefore, represent only a portion of the overall decision-making process, but these points effectively link existing planning and project development processes and practices. Many key decision points will be common among transportation agencies. Some of them are defined by law. Others have been created through the development of standard or best practice application. The individual work activities that link and feed key decision points can be quite different from country to country.

The Capacity Project C01 "A Framework for Collaborative Decision Making on Additions to Highway Capacity" of the Strategic Highway Research Program (SHRP 2, 2010) has produced the Collaborative Decision-Making Framework (CDMF) to identify key decision points (KDPs) in four phases of transportation decision-making processes. For the purpose of this project the environmental review process is considered to be merged with the permitting process. The four phases of the CDMF are:

1. Long-Range Transportation Planning
2. Corridor Planning
3. Programming

4.  Environmental Review and Permitting

The CDMF incorporates overall context sensitive solutions and project management principles and is built on a set of design goals established by the Technical Coordinating Committee for Capacity research. The design goals provide the following guidance:

- Establish a collaborative decision-making approach that identifies participant roles and responsibilities at each KDP and includes:
  - Early and on-going involvement of formal decision makers and individuals who have the potential to significantly impact the timely and cost-effective delivery of transportation improvements.
  - A tiered decision-making approach to capacity improvements that encourages binding decisions at the earliest possible point.
- Encourage timely and cost-effective project delivery through a process that:
  - Ensures transfer of information and decisions between phases
  - Encourages early and comprehensive agreement on data sources, level of detail, evaluation criteria and performance measures
  - Establishes a comprehensive and proactive risk management strategy
- Encourage a decision-making approach that evaluates transportation needs within broader community and natural contexts and integrates land planning and development policy, capital improvement planning, protection and enhancement of the human and natural environment, and addresses sustainability issues to the greatest extent possible in order to support community vision and goals.
- Encourage consideration of a wide range of options to address capacity problems during the planning phase of decision making as well as early and on-going incorporation of operational elements as a part of the overall decision-making approach.
- Establish a decision-making approach based on fulfilling the intent of legal and regulatory requirements while providing implementation flexibility and adaptability consistent with the design goals.

The CDMF is intended to be readily available to all practitioners who wish to incorporate a collaborative decision-making approach throughout the entire transportation process or only in specific areas. For this reason the ultimate vision is for the framework to be accessed through a web-based tool. The architecture of the CDMF is being designed with this in mind. The structure of the CDMF represents a series of portals through which increasingly detailed information can be retrieved for each KDP, first at the Entry Level and then at the Practitioner Level.

The diagram in Figure A1 represents the CDMF Entry Level through a series of portals in each phase of the transportation process where one or more KDPs may occur. This level demonstrates the upper-level steps in decision making as well as how the individual phases relate to one another. The community visioning process illustrated here is recommended as a best practice to ensure that the transportation decision-making process includes the larger goals and visions of the region. However, this process exists outside the transportation process, and therefore is not detailed within the CDMF. The Entry Level allows the practitioner to select an area of specific interest within the process to approach at the more detailed level.

Although the Entry Level provides a concise overview of the CDMF, transportation practitioners will need specific information at each KDP in order to consider

implementation of the collaborative decision-making process. The CDMF Practitioner Level (Figure A2) provides access to the full extent of information available at each KDP including:

- Purpose and outcome of the KDP
- Decisions made at this step
- Roles and responsibilities of the formal decision makers
- Stakeholder/project champion roles and relationships
- Supportive data, tools, and technology
- Related influencing and sub-processes
- Primary products of this step
- Associated best practices
- Linkage to other Strategic Highway Research Program (SHRP2) Capacity research such as the C02 Performance Measurement Framework

There are other community planning processes that are external to the transportation process but which have an impact on transportation decision making. Within the CDMF these processes are identified as sub-processes or influencing processes. While sub-processes have a direct effect on the transportation process through certain critical-path steps, other external processes strongly influence transportation decision making and best practice in collaboration would engage these processes as well. The CDMF contains KDPs that link the air quality, land use, and fiscal constraint sub-processes to the transportation process as well as to detailed information to allow integration of the influencing process such as the natural and human environment, safety and security planning, and capital improvement planning.

We propose that SCOR and CDMF frameworks can be integrated. Such integration would allow to assist the logistics operation planning and control with information regarding the planning and operations of the transportation system for the given infrastructure available in each case. At the same time, the actual and foreseen requirements for transportation infrastructure would be adequately considered in government and private investments.

## 4. Facing the challenge of data availability for supply chain planning and control

Supply chain planning and control processes, even though they might be designed and organized according to reference frameworks, will perform better if they use environmental information available from different sources, for instance blogs, microblogs such as Twitter, or other sources of unstructured data. Based on the proposition by Trkman et al. (2010) for supporting the demand planning and build better forecasts, we propose to use information available on the web for assisting the planning and control supply chain decision making processes. The proposed architecture is shown in Figure 3.
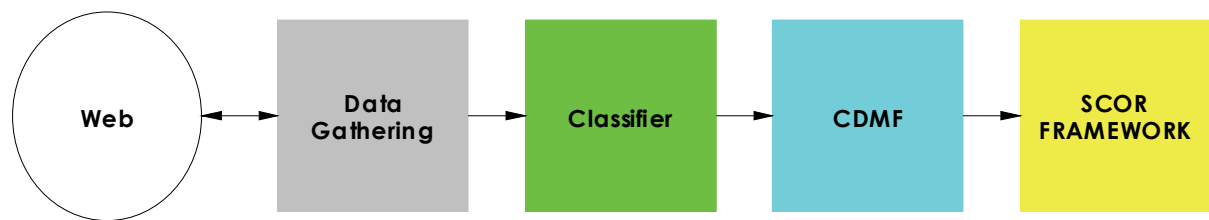


Fig. 3. Proposed architecture.

### 4.1 Automatic text classification

The processing of automatic text classification consists in selecting attributes representing the most of every class. The attributes allow to distinguish specific texts from the rest, process that permit us to assign the text being classified to a category or set of categories to which it belongs. It seeks to approximate a classifier function f, as well as possible on the basis of experience or data available through the construction of a function f 'f match most of the domain, ie:

$$f{:}D{\times}C \rightarrow \{0,1\} \tag{1}$$

$$f'{:}D{\times}C \rightarrow \{0,1\} \tag{2}$$

where $C = \{c_1, ..., c_k\}$ is the set of predefined categories and $D = \{d_1, ..., d_n\}$ a collection of documents.

### 4.2 Automatic classification types

There are a large amount of digital texts available on the web, as well as organizational databases containing useful information for a wide variety of purposes. It is expected that the amount of digital texts increase in the future so there is a need to develop machine learning techniques to analyze them effectively and efficiently. One type of analysis is the classification or pattern recognition (pattern recognition), which seeks to automatically detect certain regularities in the data using algorithms, to subsequently use these regularities to classify new data into different categories (Bishop, 2006). That is, it seeks to develop a model that, based on an analysis of the characteristics of a set of objects previously labeled, allow the assignment of correct labels to new objects.

In particular, automatic text classification is defined as assigning a Boolean value to each pair $<d_j, c_i> \in D$ x $C$, where D is the domain of texts and $C = \{c_1, ... c_{|C|}\}$ is the set of predefined tags (Sebastiani, 2002). Binary classification is the simplest and most widely used in such classification, each document is classified in one of two mutually exclusive classes, i.e. each label represents a partition or set disjoint. Moreover, multi-class classification allows each document is classified in one of several classes, also mutually exclusive. Binary classification can be extended to solve multi-class problems.

Also, a text can be classified according to one label (single-label classification) or more than one label at a time (multi-label classification). To address multi-label classification problems, two approaches are mainly used: adaptation algorithms and problem transformation (Tsoumakas, 2010). Another type of classification is based on how the assignment to classes is made: hard or soft classification (Qi & Davison, 2009). Based on the relations between classes, there are two types of classification: flat and hierarchical (Sun, et al. 2002).

### 4.3 Problem representation

The performance of artificial systems depends crucially on the quality of representation of the problem. The same task can be easy or difficult depending on how you described (Fink, 2001). When using an explicit representation of information or restrictions yield a better machine. In addition, a more complex representation can work better with simpler algorithms.

The complexity of a problem in pattern recognition, or classification, is determined by its representation in feature space (Duin & Pękalska, 2009). In particular, the representation of

the texts has a high impact on the performance of the classification task (Keikha et al., 2008). Some features of the text used to represent documents are: N-grams, words, phrases, logical terms and statements. The vector space model is one of the most used models for information retrieval, mainly due to its conceptual simplicity and the use of metaphor underlying the use of spatial proximity to estimate the semantic proximity (Manning and Schütze, 1999). This model assigns a weight to each feature of the document so that similar documents will have similar characteristics. To solve the problem of how to weigh the terms in the vector space model uses the frequency of a word in a document. However, there are effective methods for weighting terms. The basic information used in the balancing of weight is the term frequency and document frequency.

According Lan et al. (2009), two important decisions for the choice of a representation based on vector space model (VSM), are: 1) What should be the features to represent? For example, development of sub-word, word, several words or meaning, and 2) What is the weight of each feature? For example, weights can be binary, or tf-idf (Salton & Buckley, 1988).

## 4.4 Support vector machines

The theory of Support Vector Machines (SVM) is a classification technique and is based on the idea of structural risk minimization (Vapnik, 1989). In many applications, SVMs have shown a great performance, rather than traditional learning machines such as neural networks and have been introduced as powerful tools for solving classification problems. A first SVM maps the entry points to a feature space of higher dimension and finds a hyperplane that separates them and maximize the margin $m$ between the classes in this space.

Maximizing the margin $m$ is a quadratic programming problem (QP) and it can be solved by solving its dual problem by using Lagrange multipliers. Without any knowledge of the mapping, the SVM finds the optimal hyperplane using the dot product functions in the space of characteristics that are called kernels. The solution to the optimal hyperplane can be written as the combination of a few entry points are called support vectors.
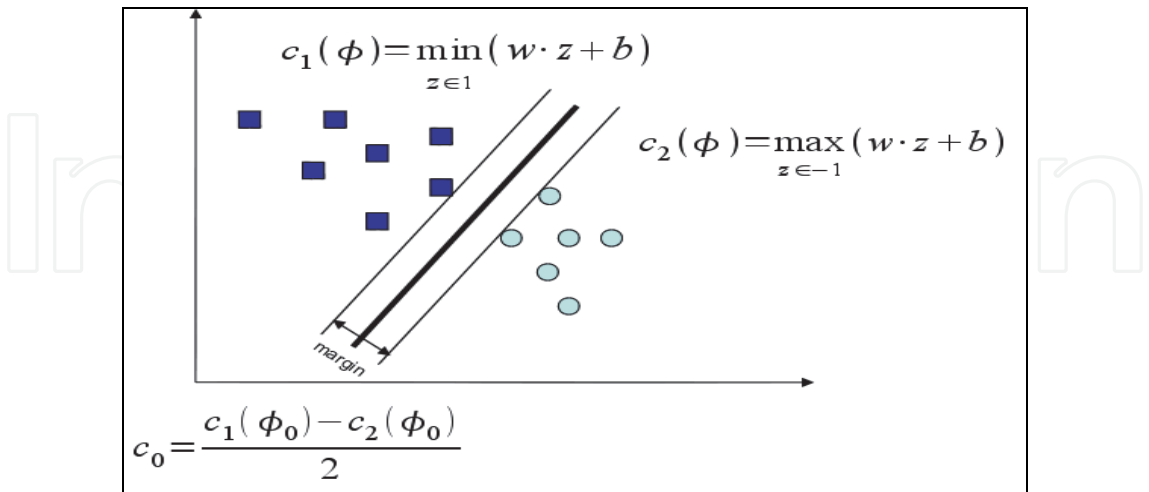


$$c_1(\phi) = \min_{z \in 1}(w \cdot z + b)$$

$$c_2(\phi) = \max_{z \in -1}(w \cdot z + b)$$

$$c_0 = \frac{c_1(\phi_0) - c_2(\phi_0)}{2}$$

Fig. 4. Separations SVM hyperplane and margins

For linearly separable, given a set $S$ of a labeled training example $(x_1, y_1), \ldots (y_i, x_i)$, each training example $x_i \in R^n$ belongs to one of two classes and has a label $y_i \in \{-1, 1\}$ for $i = 1, \ldots, l$. In most cases, the search for a suitable hyperplane in an input space is too restrictive

to be of practical use. One solution to this situation is to map the input space into a feature space of higher dimension and find the optimal hyperplane there. Let $z = \Phi(x)$ the corresponding vector notation in the feature space $Z$. Being $w$, a normal vector (perpendicular to the hyperplane), we find the hyperplane $w \times z + b = 0$, defined by the pair $(w,b)$ such that we can separate the point $x_i$ according to the $f(x_i) = \text{sign}(w \times z_i + b)$, subject to: $y_i (w \times z_i + b) \geq 0$.

In the case that the examples are not linearly separable, a variable penalty can be introduced into the objective function for mislabeled examples, obtaining an objective function $f(x_i) = \text{sign}(w \times z_i + b)$, subject to: $y_i (w \times z_i + b) \geq 1 - \xi_i$.

SVM formulations discussed so far require positive and negative examples can be separated linearly, i.e., the decision limit should be a hyperplane. However, for many data set of real life, the decision limits are not linear. To cope with linearly non-separable data, the same formulation and solution technique for the linear case are still in use. Just transform your data into the original space to another space (usually a much higher dimensional space) for a linear decision boundary can separate positive and negative examples in the transformed space, which is called "feature space." The original data space is called the "input space." Thus, the basic idea is that the map data in the input space $X$ to a feature space $F$ via a nonlinear mapping $\Phi$,

$$\Phi: X \to F \qquad (3)$$

$$X \to \Phi(x) \qquad (4)$$

The problem with this approach is the computational power required to transform the input data explicitly to a feature space. The number of dimensions in the feature space can be enormous. However, with some useful transformations, a reasonable number of attributes in the input space can be achieved.

Fortunately, explicit transformations can be avoided if we realize that the dual representation, both the construction of the optimal hyperplane in $F$ and the corresponding function assessment decision/classification, only requires the evaluation of the scalar product $\Phi(x) \cdot \Phi(z)$ and the vector $\Phi(x)$ is never allocated in its explicit form. This is a crucial point. Thus, we have a way to calculate the dot product $\Phi(x) \cdot \Phi(z)$ in the feature space $F$ using the input vectors $xyz$, then it would not need to know the feature vector $\Phi(x)$ or even mapping function $\Phi$. In SVM, it's done through the use of "kernel function", which is referred to as $K$. $K(x,z)$ equals to $\Phi(x) \cdot \Phi(z)$ and are exactly the functions for calculating dot products in the transformed feature space with input vectors $x$ and $z$. An example of a kernel function is the polynomial kernel, $K(x,z) = \langle \mathbf{x},\mathbf{z} \rangle^d$, which can replace all dot products $\Phi(x) \cdot \Phi(z)$. This strategy of directly using a kernel function to replace the dot products in the feature space is called "kernel trick." Where would never have to explicitly know what function $\Phi$ is. However, the question remains how to know a kernel function without making its explicit referral. That is, ensuring that the kernel function is actually represented by the dot product of the feature space. This question is answered by the Mercer's Theorem (Cristianini & Shawe-Taylor, 2000).

### 4.5 Automatic classification of opinion (sentiment analysis)

Today, large amounts of information are available online documents. In an effort to better organize the information for users, researchers have been actively working the problem of automatic text categorization. Most of this work has focused on the categorization of

categories, trying to sort the documents according to subject (Holts et al., 2010). However, recent years have grown rapidly in online discussion groups and sites reviews, where a crucial feature of the articles published is his way or global opinion on the subject, for example if a product review spoke positively or negatively (Pang & Lee, 2008). The labeling of these items with your sentiment would provide added value to readers, in fact, these labels are part of the appeal and added value of sites like www.rottentomatoes.com, which labeled the movie that do not contain explicit rating indicators and normalizes the different rating systems that guide respondents' sense. It would also be useful in business intelligence applications and recommender systems, where user input and feedback can be quickly summarized. On the other hand, there are also potential applications for filtering messages, for example, one might be able to use the information to recognize the meaning and discard comments that were not interested in reading. This chapter examines the effectiveness of applying machine learning techniques for the classification problem of meaning. A challenging aspect of this problem that seems to distinguish it from the traditional classification based on themes is that although the topics are often identified by keywords, the meaning can be expressed more subtly.

An expert system using machine learning for text categorization has a relatively poor performance compared to other automatic classification applications. Moreover, differentiating positive from negative text comments is relatively easy for humans, especially when comparing to the problem of standard text categorization, where issues can be closely related. There are people whose use specific terms to express strong feelings, so it might be sufficient to generate a list of terms to classify the texts. Many studies indicate that it is worth to explore techniques based on domain-specific corpus, instead of relying on prior knowledge to select the features for feelings and sorting.

## 5. Case of study: Premium Chilean wine supply chain

For testing the supply chain framework and its assisting information retrieval technology, we select model the premium Chilean wine supply chain and use Twitter available comments as unstructured data source for assisting the demand planning and the supply chain control. This domain is experimentally convenient because there are large collections online readily available, but they are not labeled. Therefore, there is a need for hand-label data for supervised learning. The comments were taken automatically from the popular Twitter platform and categorized into one of three categories in relation to demand growth: positive, negative, or neutral. For the situation at hand, we assume that an increment of positive comments implies that demand will increase (at least for the next business cycle). While neutral comments are considered as not affecting the demand. Comments considered as advertisement where classify within this category. Finally, negative comments are considered to affect the demand negatively.

Chile has a long history in winemaking (Visser, 2004). In 1551, a Spanish conqueror managed to make wine at a location 500 kilometers north of Santiago. During the colonial period, wine was made for religious purposes. In the 18th and 19th century, rich families in Chile made wine imitating French Chateaux and thus importing classical grape varieties and technology from France. The outbreak of Phylloxera in Europe at the end of the 19th century stimulated the export of quality wines. In the 20th century, wine production slowed down, as import-substitution policies did not favor exports and wine-makers depended on a small domestic market. In the 1980s, changes in macroeconomic policies and national law

joined crucial developments in the domestic and international wine markets, boosting vineyard area, wine production and exports in the 1980s and the 1990s.

It takes about three years before new vines are in production, so the growth of wine production is likely to increase at least until 2004, as a result of the accelerating increase of the planted area in 1999/2000. In international perspective, only China and Australia surpass Chile regarding the speed of increase in the vineyard area during 1995-2000, with a 57 and 73% respectively.

Despite the fast increase of the vineyard area after 1995, Chile ranks 11th in the world on this count (ibid.), holding a share of 1.3% in 2001. Spain is first on the list, with a 15.5% share of the global vineyard area. France (11.9%), Italy (11.5%), Turkey (6.7%), and USA (5.2%) follow, while Argentina had a 2.6 % share in 2001.

The industry's main focus is red vines. Important grape varieties are Cabernet Sauvignon and Merlot. Syrah and Carmenère are relatively new additions to Chilean wine. The planted area of these four wine grape varieties increased considerably. The Carmenère grapes will continue to increase in importance during the following years, as this variety disappeared in Europe (where it comes from), due to the world wars and several plagues. At the moment, Chilean wine producers aim at expanding Carmenère production, branding it as a typical Chilean vine, like Shiraz reds for Australia or Malbec for Argentina.

Chile's wine industry is an example of an effective turnaround from a focus on domestic towards export markets. Several indicators can be used to sustain this point, e.g. the share of wine sold abroad; export sales volume, value, and share in global markets; the geographical diversification and penetration of markets; and the number and location of exporting firms. The share of Chilean wines sold abroad increased from 7% in 1989 to 63% in 2002. In volume terms, only 8,000 hectoliters were exported in 1984, a figure rising to 185 thousand in 1988, and then accelerating throughout the 1990s, so that in 2002, more than 3.5 million hectoliters of Chilean wine found their way to the world market. This is the fastest growth recorded for New World wine producers during the period under review (Coelho 2003). With this, Chile's share in global wine export volume rose from about zero in 1984 to over 4% in 2000. Export value rose from a meager 10 million US-dollars (FOB) in 1984, to 145 million US-dollars (FOB) in 1994 and a dazzling 602 million US-dollars (FOB) in 2002. Premium Chilean wine supply chain considers national and international suppliers as well as mostly international customers (Figure 5).

According to the architecture proposed and shown in Figure 3, a total of 1004 Twitter comments were gathered from January 26, 2011 until March 29, 2011. An example of twitts comments are shown in Table 1.

Then, a manual classification was performed on a subset of 200 comments, to label them into positive, negative, or neutral categories, in order to use them as testing and training sets to be input to the Support Vector Machine devised. The results of the classification process performed over the entire data set are shown in Table 2.

Given the result in Table 2, the behavior of the demand must be expected to grow. How much growing in the demand should be expected is matter of a business intelligence system. These scattered signals gathered in the system we propose, must act jointly with systems at every level in the logistics chain to prepare each company for the situation ahead. According to our solution schema, this information should be passed through the highway capacity framework to the SCOR supply chain model and plan accordingly. Action regarding selection of transportation routes and modes as well as production, supply, and logistics processes planning in the supply chain should take place after feedback information is

obtained. Long term planning must take place based on aggregated information, both from structured and unstructured information.
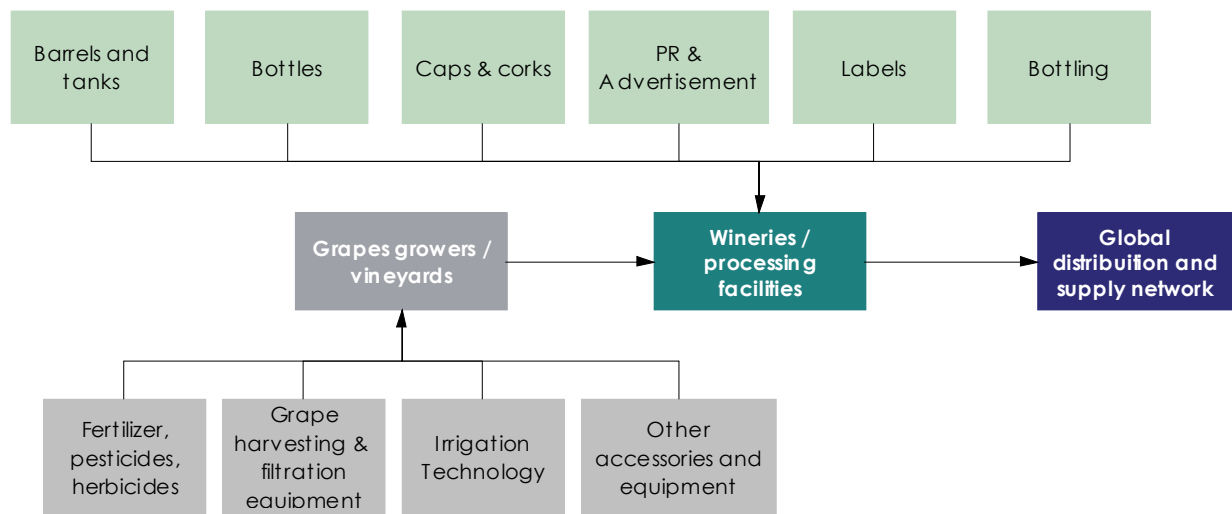


Fig. 5. Premium Chilean wine supply chain.

| Date | Comment | Category |
|------|---------|----------|
| 01/26/11 10:47 PM | Tabali Reserva Especial 2008 Syrah http://bit.ly/gdKos3 | neutral |
| 02/02/11 10:52 PM | So Jr. wants to do study abroad in Chile next year. My 1st question is..."How much wine can you bring back home?" Me loves Chilean wine. | positive |
| 02/04/11 03:32 PM | Jeez, you could clean windows with these personalized bottles of chemically-enhanced Chilean wine. | negative |
| 02/10/11 03:50 PM | Enjoying a Chilean wine this Valentine's Day? Whether it's red, white, sparkling or still, we want to hear about it! | positive |

Table 1. Examples of twitts about "Chilean wine"

|  | Neutral | Positive | Negative |
|------|---------|----------|----------|
| Accuracy | 19.64% | 95.71% | NA |
| Percentage | 38% | 60% | 2% |

Table 2. Performance measurements of sentiment classificator.

## 6. Conclusion

An integrated framework based on SCOR and CDMF by the U.S. Transportation Research Board for modeling supply chains is proposed. The proposed framework is comprehensive in terms of considering all the processes taking place in the supply chain for a given product and at the same time assist by taking into account the transportation system capacity. We also propose the operation of the supply chain model, obtained with the integrated

framework, should operate considering both structured data (available mostly in companies or government agencies databases) and unstructured data (available from web sources such as social networks). However, the enrichment that unstructured data provides to classical decision making processes is important but does not eliminates the need for structured data. Nevertheless, the amount of unstructured data available on the web is increasing by the minute and its processing requires of powerful technologies of data processing and storage, becoming available in a continuous basis. Thus, the processing of huge amounts of, apparently, unrelated data produces rich information at low price, situation that has no comparison to structured data (or that might be obtained at a very high price). The proposed integrated framework and information retrieval assisting technology is scalable to supply chains and applications in fields other than logistics.

## 8. Appendix
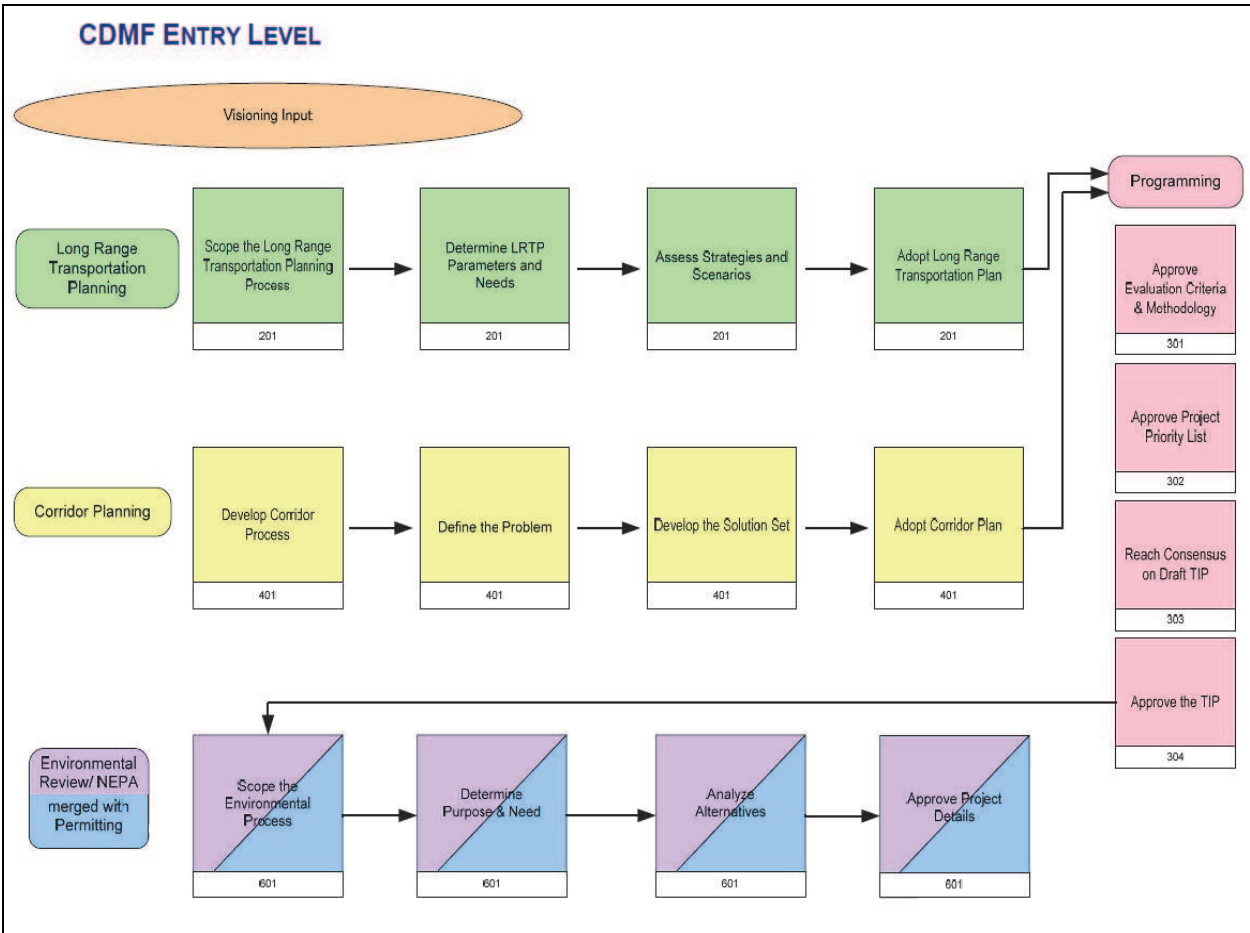


Fig. A1. Collaborative Decision-Making Framework Entry Level (SHRP 2, 2010)
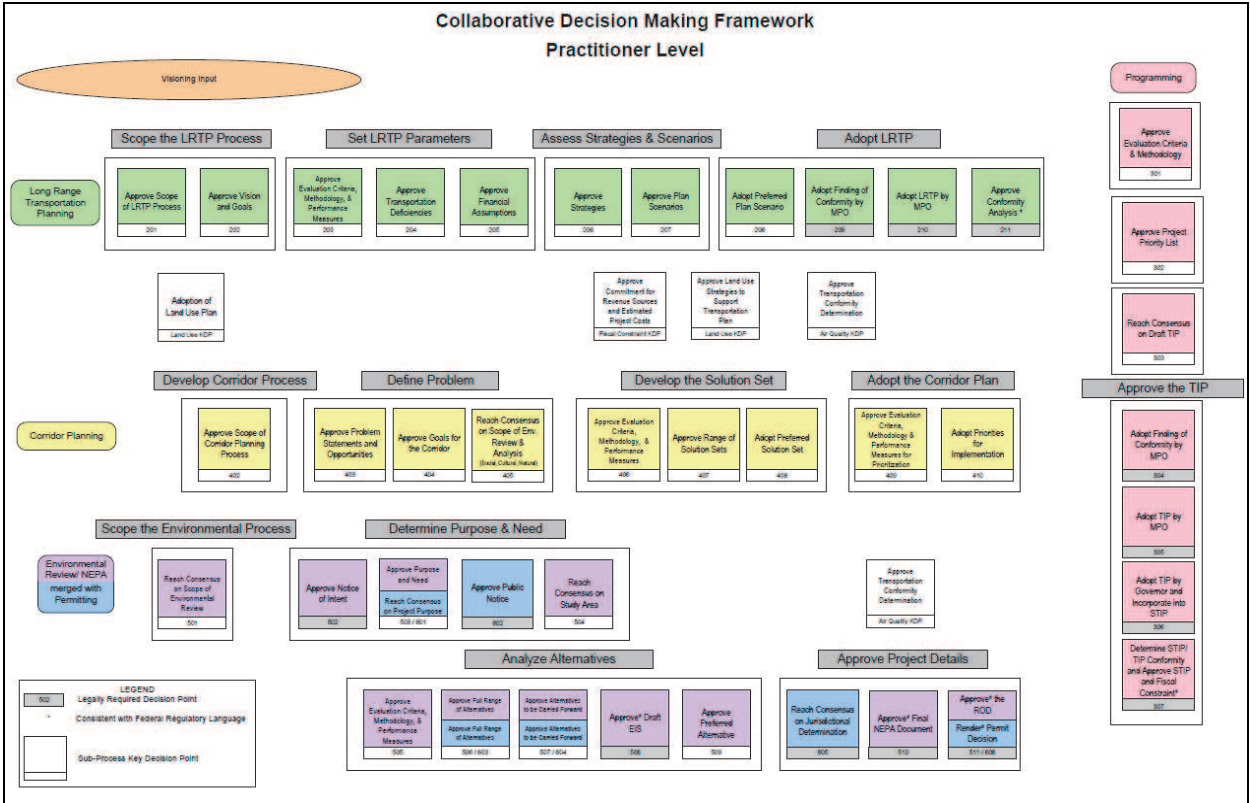
Fig. A2. Collaborative Decision-Making Framework Practitioner Level (SHRP 2, 2010)

## 9. References

Bishop C. (2006). *Pattern Recognition and Machine Learning*. Springer.

Coelho, A. (2003), *Presentation at an EADI workshop on Clusters and Global Value Chains in the North and the Third World*, organized at the Università del Piemonte Orientale, Novara, Italy, October 30-31.

Corsten, H., 2001. *Einführung in das Supply Chain Management*. R. Oldenbourg Verlag, München.

Cristianini, N. & Shawe-Taylor, J. (2000). *Support Vector Machines and other kernel-based learning methods,* Cambridge University Press, 2000.

Duin R. P. W. & Pękalska E. (2006) Object Representation, Sample Size, and Data Set Complexity, Data Complexity in Pattern Recognition, J. Lakhmi, W. Xindong (Ed.), Springer London.

Fink E. (2001) Automatic evaluation and selection of problem-solving methods: Theory and experiments, *Journal of Experimental and Theoretical Artificial Intelligence* 16(2) (2004), pp. 73-105.

Holts, A., Riquelme, C. & Alfaro, R. (2010) Automated Text Binary Classification Using Machine Learning Approach, Proceedings of the Chilean Society of Computer Science Conference (SCCC), Antofagasta, November 2010, pp. 212-217.

Hvolby, H. & Trienekens, J. (2010). Challenges in business systems integration. *Computers in Industry*, 61 (August 2010), pp. 808–812

Keikha M., Razavian N.Sh., Oroumchian F. & Razi H. S. (2008) Document representation and quality of text: An analysis. In *Survey of Text Mining II: Clustering, Classification, and Retrieval*, Springer-Verlag, London, pp. 135-168.

Lan M., Tan Ch. L., Su J. & Lu Y. (2009), Supervised and traditional term weighting methods for automatic text categorization, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31, 721-735.

Manning Ch. & Schütze H. (1999), *Foundations of statistical natural language processing*, The MIT Press.

NSTPRSC. 2008. Transportation for Tomorrow. National Surface Transportation Policy and Revenue Study Commission, *Transportation Research Board*. (2008). Available from http://transportationfortomorrow.com/final_report/technical_issue_ papers.htm

Pang B. & Lee L. (2008), Opinion Mining and Sentiment Analysis, *Foundations and Trends in Information Retrieval*, v.2 n.1-2, pp. 1-135.

Qi X. & Davison B. D. (2009) Web page classification: features and algorithms, *ACM Computing Surveys*, vol 41 N°2, Article 12.

Röder, A. & Tibken, B (2006). A methodology for modeling inter-company supply chains and for evaluating a method of integrated product and process documentation. *European Journal of Operational Research*, 169 (April 2005), pp. 1010–1029

Salton G. & Buckley Ch. (1988) Term-weighting approaches in automatic text retrieval, *Information Processing and Management: an International Journal* 24, no. 5, pp. 513-523.

Schönsleben, P., 2000. Integrales Logistikmanagement—P*lanung und Steuerung von umfassenden Geschäftsprozessen*. Springer-Verlag, Berlin.

Sebastiani F. (2002) Machine learning in automated text categorization, *ACM Comput. Surveys* 34, no. 1, pp. 1-47.

SHRP 2, 2010. Performance Measurement Framework for Highway Capacity Decision-Making, Strategic Highway Research Program 2, Transportation Research Board. (2010). Available from http://www.trb.org

Sun A., Lim E. P. & Ng W. K. (2002) Web classification using support vector machine. *In Proceedings of the 4th International Workshop on Web Information and Data Management (WIDM)*. ACM Press, New York, NY, pp. 96–99.

Supply Chain Council, 2010. *Supply-Chain Operations Reference-model—Overview of SCOR Version 10.0*, Pittsburgh. Available from: http://www.supply-chain.org

Trkman, P, McCormack K., Oliveira M. P. V. & Ladeira M. B., (2010) The Impact of Business Analytics on Supply Chain Performance. *Decision Support Systems*, Vol. 49, No. 3, pp. 318–327.

Tsoumakas G., Katakis I. & Vlahavas I. (2010), *Mining multi-label data, Data Mining and Knowledge Discovery Handbook*, 2nd edition, O. Maimon, L. Rokach (Ed.), Springer.

Vapnik, V. N. (1989). Statistical Learning Theory. Wiley-Interscience.

Visser, E. 2004. *A Chilean wine cluster? Governance and upgrading in the phase of internationalization*. (September 2004). ECLAC/GTZ project on "Natural Resource Based Strategies Development" (GER 99/128)

Williams, T.J. (1992). The Purdue Enterprise Reference Architecture, Instrument Society of America, Research Triangle Park, USA, 1992.

**Supply Chain Management - New Perspectives**

Edited by Prof. Sanda Renko

Over the past few decades the rapid spread of information and knowledge, the increasing expectations of customers and stakeholders, intensified competition, and searching for superior performance and low costs at the same time have made supply chain a critical management area. Since supply chain is the network of organizations that are involved in moving materials, documents and information through on their journey from initial suppliers to final customers, it encompasses a number of key flows: physical flow of materials, flows of information, and tangible and intangible resources which enable supply chain members to operate effectively. This book gives an up-to-date view of supply chain, emphasizing current trends and developments in the area of supply chain management.

**How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

# INTECH
open science | open minds