CarbonDB: a Semantic Life Cycle Inventory Database

Benjamin Bertin
Université de Lyon, CNRS
INSA-Lyon, LIRIS, France
bbertin@liris.cnrs.fr

Vasile-Marian Scuturici
Université de Lyon, CNRS
INSA-Lyon, LIRIS, France
mscuturici@liris.cnrs.fr

Jean-Marie Pinon
Université de Lyon, CNRS
INSA-Lyon, LIRIS, France
jmpinon@liris.cnrs.fr

Emmanuel Risler
Université de Lyon, CNRS
INSA-Lyon, ICJ, France
emmanuel.risler@insa-lyon.fr

ABSTRACT

We demonstrate CarbonDB, a web application for Life Cycle Inventory data management. Life Cycle Assessment provides a well-accepted methodology for modelling environmental impacts of human activities. This methodology relies on the decomposition of a studied system into interdependent processes in a phase called Life Cycle Inventory. Several organisations provide processes databases containing thousands of processes with their interdependency links. The usual workflow to manage those databases is based on the manipulation of individual processes, which turns out to be a very harnessing work even if there are strong semantic similarities between the involved processes. In previous publications, we proposed a new workflow for LCA inventory databases maintenance based on the addition of semantic information to the processes they contained. This method considerably eases the modeling process and offers a synthetic view of the dependencies links. We created a web application based on this approach composed of a back-end for data management and a front-end for searching processes and visualize the dependencies links in a graph.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Clustering; H.3.2 [**Information Storage**]: Record classification; H.3.1 [**Content Analysis and Indexing**]: Indexing methods; H.2.8 [**Database Management**]: Scientific Databases

Keywords

Clustering, Environmental database, Life Cycle Assessment, Onotology, Semantic annotation

1. INTRODUCTION

In order to reduce the environmental impact of human activities, it is necessary to model and evaluate the environmental effects of these activities. This is the objective of the *Life Cycle Assessment* (LCA) method [3], which aims at determining the environmental impacts of a product, a service or, generally speaking, any human activity. This method can take into account all the life cycle stages of a product such as manufacture, use and recycling. LCA can assess various environmental impacts, like greenhouse gases emissions or chemical products dissemination.

CIKM'12, October 29–November 2, 2012, Maui, HI, USA.
ACM 978-1-4503-1156-4/12/10.

During an LCA study we factorize a studied system into interrelated elementary processes associated to environmental impacts in order to achieve a *Life Cycle Inventory* (LCI). Elementary processes are related to specific stages of a life cycle or to any human activities (e.g., energy production, air plane trips, etc.). These processes can depend on other processes, for instance: car production depends on steel production. The accepted methodology for the LCI is to use an Input/Output (I/O) matrix [10] to model inter-process interactions or interactions between processes and the environment.

Several agencies and companies provide Life Cycle Inventories databases (like [6][1][2]) that are used by LCA practitioners to do an LCA study. But those databases can contain thousands of processes linked together. The model is therefore difficult to understand unless the practitioners do an in-depth analysis. Yet, semantic similarities are noticeable in LCI databases: the dependency links between processes in LCI databases contain many network constructions that are both topologically and semantically close. For instance, in order to produce electricity from coal it is necessary to transport it. In an LCI database, we find several processes corresponding to electricity production from coal for every type of coal that can be used to produce electricity (lignite coal, bituminous coal or sub-bituminous coal). These processes are all linked to several merchandise transportation processes depending on the transportation systems (the coal can be transported using a truck, a barge, etc.).

But, if the dependency links are scattered into an I/O matrix, it is hard to apprehend those similarities. Meanwhile, maintaining an LCI database can be tedious and harnessing if we have to update many semantically close relations.

Thus, we proposed a new methodology in [4] and [5] to model an LCI database based on the semantics of the processes. In our approach, we semantically index the processes and we use this semantic indexing to semantically regroup the processes. Then, we use those groups to model the dependency links between multiple processes. With this methodology we address the two problems we identified: we offer a bird's eye view of the database and we ease the database management using the regrouped processes.

Our model is based on the coexistence of two digraphs. The first one, called the *macro-graph*, contains dependencies between regrouped processes (or *macro-processes*). The second one, called the *detailed-graph*, is a transposition of the I/O matrix. In our approach, rather than creating dependency links between individual processes, we create links between macro-processes. Then we translate those relations

into inter-process relations in the detailed-graph.
We built a web application implementing this approach restricted to greenhouse gases emissions: CarbonDB[1]. It contains a back-end for LCI data management. It also contains a front-end offering a radial graph visualization of the macro-graph and the detailed-graph, a keyword-based search engine and an impact visualization interface.

2. OVERVIEW OF THE APPROACH

The goal of Life cycle assessment methodology is to evaluate the environmental impacts of, for instance, a product or a company's activity. This can be achieved by creating an inventory of elementary flows from and to the environment, for every step of a product's production process or for every activity of a company [8][9]. In the LCA terminology, these steps or activities are called *processes*.
A process is associated with environmental impacts (e.g., the greenhouse gases emitted or the resulting water pollution) and with other processes. For instance, it is necessary to extract and transport coal in order to run a coal power plant. The total amount of environmental impacts of a process is the sum of its own impacts (i.e., that are not from its predecessors) and the impacts of its predecessors multiplied by a scalar dependency coefficient.
The dependency links between the processes can be represented using a digraph. Let $G(V, E)$ be the digraph representing dependency links between processes; in our terminology, this graph is called *detailed-graph*. The vertices set V is the set of processes. The edges set E contain the relations between processes and the set of weights associated to the edges is the set of coefficients. Let p_1 and p_2 be vertices, then a directed edge from p_1 to p_2 means that process p_2 depends on p_1.
Determining the impacts for a specific process requires to recursively calculate the impacts of its predecessors. As explained in [12], the dependency coefficients form a basic system of linear equations. Thus it is possible to calculate the impacts of the processes using iterative methods[2] or any direct method (like the Gaussian elimination)[13].
Our approach is based on the existence of two layers of directed graphs. The first graph, called the *detailed-graph*, contains the dependency relations between processes. The second graph, called the *macro-graph*, contains the relations between groups of processes (or *macro-processes*). The macro-graph offers a simplified view of the data contained in the detailed-graph and eases the expression of new dependency relations between semantically close processes. In order to semantically regroup processes, we choose to index them with a set of keywords that are stored in an ontology [7][11]. The vocabulary of this simple ontology is composed of keywords and predicates to create binary relations between these keywords.
Using this ontology, we can regroup processes and dependency coefficients into semantic groups. A macro-process is similar to a multidimensional matrix, where each dimension is a set of keywords. The dimensions are described using a query over the ontology. Then, we can create dependency relations between macro-processes using *macro-coefficients*, in the same way that we create dependency relations between processes using individual coefficients.

[1]Because our approach is focused on modelling the dependency relations between the processes, we choose to restrict the environmental impacts to green house gases emissions, thus we named this application CarbonDB.

[2]Life Cycle Inventory data is subject to uncertainty. It is therefore possible to use an iterative method and stop the algorithm if the delta obtained between two iterations is smaller than the uncertainty. This could significantly lower the computation time compared to a direct method.

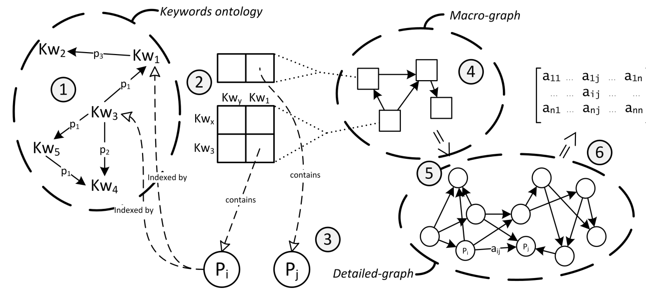

Figure 1: Our approach workflow

The methodology based on those three graphs is summarized in Figure 1. The first part contains the keyword ontology. The second part contains two macro-processes referencing some processes shown in the third part (those processes are indexed with keywords). The fourth part contains the macro-graph and is translated into the detailed-graph in the fifth part. The sixth part contains the I/O matrix extracted from the detailed-graph.
Equipped with this methodology, if we want to create a new LCI database, we have to follow six steps. During the first one we need to create the keywords ontology. In the second step we create some macro-processes using the ontology. Those macro-processes reference several processes depending on the keywords coordinates of the cells in the macro-processes. Because the database is empty, in the third step, individual processes are automatically created according to the coordinates of the macro-processes cells. Then, in the fourth step, we can create dependency relations between the macro-processes. Finally, the macro-graph is automatically translated into a detailed-graph in the fifth step and, in the sixth step, the coefficients are extracted and the impacts are calculated solving the corresponding linear equation system. This approach is described in-depth in [4] and [5].
If we add a new macro-process to an already existing LCI database, the processes it references can already exist. Hence, we would enrich the dependency network of those processes. Furthermore, as the macro-processes are defined using the ontology, we can edit macro-processes by just editing the ontology.

3. APPLICATION

We developed the CarbonDB web application using PHP and Zend Framework on the server side, and Yahoo UI and Protovis for the client interfaces. This application is composed of a back-end with interfaces to manage the database and a front-end with a search engine and a visualization interface of the two digraphs (Figure 2 shows the detailed-graph visualisation interface of the front-end).
The back-end contains several components corresponding to the different steps shown in Figure 1: 1) A simple ontology editor to manage predicates, keywords and relations between keywords; 2) An interface to create groups of processes and coefficients; The dimensions of a group are defined as inter-

section of sets of keywords[3]; A set of keywords contains all the keywords linked to a specific keyword considering only one predicate; 3) An interface to edit the macro-relations between groups of processes using groups of coefficients.
The translation of the macro-relations into a set of detailed-relations is done in batch mode. As the detailed-graph contains only generated relations, this procedure does the following actions: i) delete all the detailed-relations; ii) translate every (well defined) macro-relations; iii) calculate the impacts of the processes.
The front-end has two main components: a keyword-based search engine and a visualization interface. The latter is composed of an interactive visualization of the macro-graph and the detailed-graph, and a panel containing the details of the active node in the graph visualization. The graph visualization also serves as a navigation interface in the graphs. Both graphs are displayed in a radial way: we focus on one node and display its predecessors and successors on semicircles and their neighbours into concentric semicircles. Finally, we added a logging functionality, which offers a branching and tagging capability. Thus we can manage multiple databases, consult the user actions history and revert any change.

4. DEMONSTRATION DESCRIPTION

Back-end For the demonstration, we will show the conference participants how to create a new macro-relation, starting from the groups definitions with the keywords ontology already containing some keywords. Data used during the demonstration will come from the U.S. National Renewable Energy Laboratory (NREL) Life Cycle Inventory Database [2]. We will only use the inventory data for electricity production in the U.S. More specifically, our demonstration will be focused on the electricity production from coal, which is complex enough to apprehend the importance of the macro-graph. First, we define a group containing transport processes, a group containing electricity production from coal (there are several types of coal that can be used to produce electricity: lignite coal, bituminous coal, etc.) and a group containing the dependency coefficients between those two groups (we also set all the greenhouse gases impacts of the processes and the coefficients values). Then we can start the translation and impacts calculation processing. To demonstrate the importance of the ontology, we finally add a new keyword corresponding to a new mode of transport that could be used to transport coal. Consequently, the group containing the transport processes and the group of coefficients are automatically updated and, after setting the proper value for the newly created process and coefficients, we can start again the processing. At the end we will obtain new detailed-relations between this new transport process and the already existing electricity production from coal processes.
Front-end The results of those actions in the back-end are then shown in the front-end, where the conference participant can navigate on the macro-graph (containing only two nodes) and the corresponding detailed-graph. In order to demonstrate the keyword-based search engine and the navigation interface, we will switch to another database filled with all the processes from the NREL database for the electricity production in the U.S.

A demonstration video can be found at `http://liris.cnrs.fr/~bbertin/carbondb.mp4`.

[3]As it is sometimes necessary to define a dimension containing only one keyword, we introduce the notion of *common keyword*, behaving like a dimension and applying to the coordinates of every cell in the group.

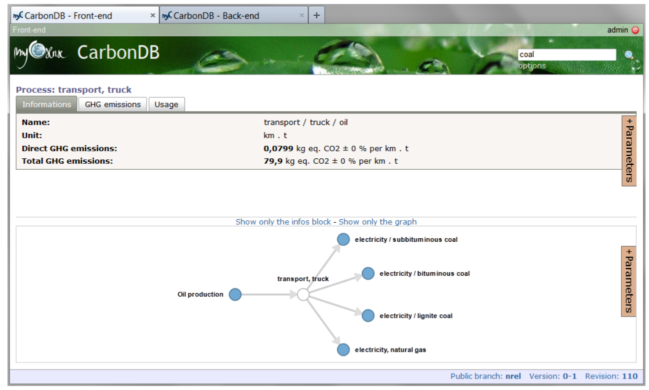

Figure 2: Capture from the front-end of CarbonDB

5. REFERENCES

[1] GABI Life Cycle Inventory Databases, `http://www.gabi-software.com/`
[2] U.S. Life Cycle Inventory Database, `http://www.nrel.gov/lci/`
[3] ISO 14044 (2006): Environmental Management – Life Cycle Assessment – requirements and guidelines. International standard, International Organisation for Standardisation (ISO), 2006.
[4] B. Bertin, M. Scuturici, J.M. Pinon, E. Risler. A Semantic Approach to Life Cycle Assessment Applied on Energy Environmental Impact Data Management *Workshop on Energy Data Management in conjunction with EDBT*, 2012.
[5] B. Bertin, M. Scuturici, J.M. Pinon, E. Risler. Semantic modelling of dependency relations between Life Cycle Analysis processes *International Conference on ICT as Key Technology against Global Warming in conjunction with DEXA*, 2012.
[6] R. Frischknecht, G Rebitzer. The ecoinvent database system: a comprehensive web-based LCA database. *Journal of Cleaner Production*, 2005.
[7] T. R. Gruber. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 1993.
[8] J. B. Guinée. *Handbook on Life Cycle Assessment: Operational Guide to the ISO Standards*. Springer, New York, 2002.
[9] R. Heijungs, S. Suh. *The computational structure of life cycle assessment*. Kluwer Academic Publishers, Dordrecht, 2002.
[10] W. Leontief. *Input-Output Analysis*, pages 53–83. University of British Columbia Press, 1986.
[11] D. L. McGuinness. *Ontologies come of age*, page 171–192, in: The Semantic Web: Why, What, and How. MIT Press, 2002.
[12] G. P. Peters. Efficient Algorithms for Life Cycle Assessment, Input-Output Analysis, and Monte-Carlo Analysis. *The International Journal of Life Cycle Assessment*, 2007.
[13] WK Nicholson. *Elementary linear algebra with applications*. PWS-Kent Publishing Company, 1990.