A Knowledge Graph Framework for Impact Calculation in Life-Cycle Assessment

Original
A Knowledge Graph Framework for Impact Calculation in Life-Cycle Assessment / Diamantini, C., Potena, D., Rossetti, C., Storti, E.. - 4002:(2025), pp. 32-38. (3rd International Workshop on Knowledge Graphs for Sustainability, KG4S 2025 Portoroz (Slovenia) June 1st, 2025).

Publisher:
CEUR-WS

Published
DOI:

A Knowledge Graph Framework for Impact Calculation in Life-Cycle Assessment

Claudia Diamantini[1], Domenico Potena[1], Cristina Rossetti[1,2] and Emanuele Storti[1,*]

[1]DII, Università Politecnica delle Marche, via Brecce Bianche, Ancona, 60131, Italy
[2]DAUIN, Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino, 10129, Italy

Abstract

Sustainability assessments are increasingly critical for evaluating the environmental impacts of business activities. Life-cycle assessment (LCA) is a key methodology in measuring these impacts, but integrating and analyzing data from diverse enterprise data sources to compute LCA indicators remains a challenging task. In this paper, we propose an approach that leverages Knowledge Graphs to formally model LCA indicators and their associated mathematical calculation formulas. The graph serves as a flexible schema supporting enterprises in documentation and mathematical interpretation of indicators for LCA, by linking their internal data sources to the Graph. On top of it, a suite of reasoning-based services is presented to automate the calculation of indicators from available data sources through algebraic manipulation, and to facilitate collaboration among a network of enterprises by enabling consistent comparison of sustainability assessments.

Keywords

Knowledge Graph, Life-cycle assessment, Performance Indicator, KPI, Sustainability

1. Introduction

Sustainability evaluation is an analytical process aiming to understand, measure and improve the impact of human activities on the environment, society, and the economy. This approach is not limited to document the results of a project, but serves as a strategic guide to steer decisions towards a balance between economical progress and environmental protection. In recent years, this concept has become central to corporate planning and public policies. Organizations use sustainability evaluation to demonstrate transparency, identify new opportunities for innovation and manage risks related to climate change, scarcity of natural resources and increasing expectations of social responsibility.

Among the most established tools for sustainability evaluation, Life Cycle Assessment (LCA) [1] is a standardized methodology for the assessment of environmental performances of an activity. The main goal is to evaluate the potential environmental impacts considering all the life cycle phases related to the activity, from the acquisition of raw materials to the final disposal, providing valuable information for developing improvement interventions, such as optimizing logistics or adopting low-impact fuels. LCA involves four main steps: (1) goal and scope definition; (2) life cycle inventory (LCI), consisting in listing and quantifying input and output data of the activity; (3) life cycle impact assessment (LCIA), which provides evaluations on potential environmental impacts; and (4) interpretation of results. In details, the LCIA consists of grouping the different emissions and used resources based on impact categories (e.g., 'climate change', 'human toxicity', 'water use') and then convert them to common units allowing comparison [2].

Several impact assessment methods exist for different impact categories, such as Environmental Footprint (EF)[1]. For each impact category there can be multiple indicators which measure specific environmental performance. When an organization wants to assess its LCIA score across the entire

---

The 3rd international Workshop on Knowledge Graphs for Sustainability - KG4S, ESWC 2025, June 1–5, 2025, Portoroz, Slovenia

[*] Corresponding author.

c.diamantini@univpm.it (C. Diamantini); d.potena@univpm.it (D. Potena); cristina.rossetti@polito.it (C. Rossetti); e.storti@univpm.it (E. Storti)
0000-0001-8143-7615 (C. Diamantini); 0000-0002-7067-5463 (D. Potena); 0009-0003-5243-4249 (C. Rossetti); 0000-0001-5966-6921 (E. Storti)

[1] https://green-business.ec.europa.eu/environmental-footprint-methods/life-cycle-assessment-ef-methods_en

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

supply chain, it has to evaluate different impact indicators for each activity according to the impact category involved. This process requires information about products, emissions, pollution, waste on each activity, thus making the LCA a data-intensive procedure.

Many databases collecting LCI data are available with the aim of supporting sustainability assessment within organizations. Among these, one of the most relevant is Ecoinvent[2], which collects global data containing detailed information on human activities and related environmental impacts. Ecoinvent datasets also store mathematical relationships on different fields of an activity to calculate quantitative measures, such as the amount of a certain emission. These formulas can serve as validation and support for the calculation of certain measures, avoiding errors in the final assessment of impacts. In this context, an important challenge concerns the lack of formalization of LCA concepts and standardization across LCI databases. In fact, most LCI databases are based on traditional RDBMS systems, without an explicit semantic support, thus rising interoperability and scalability issues [3]. In order to overcome these challenges, some works have focused on employing semantic technologies, such as ontologies and Knowledge Graph (KG), to represent the semantics of LCI data, thus providing a common and shared knowledge base for LCA analysis. Among them, the KG in [4] is aimed to enhance automated LCA, while in [3] it supports integration and storage of Ecoinvent data. In [5], an ontology is used for documentation purposes and is further extended in [6] to build a semantic catalog for LCA data, and in [7] to support Life Cycle Sustainability Assessment (LCSA). Despite these efforts, most works focus on defining of the main concepts involved in LCIA such as activities and flows, while omitting the explicit representation of indicators, analytical relationships among them and calculation methods. This omission is crucial when LCA practitioners need to customize data for specific activities, like the quantity of raw materials. In fact, LCI databases such as Ecoinvent provide quantitative data based on estimates made on a global or local scale. This kind of information is not appropriate for specific cases where, for example, an organization wants to input its own production and resource use data to ensure a more accurate impact assessment.

In this paper, we address the mentioned issues by proposing a semantic-based approach to support accurate enterprise-tailored assessment of environmental impact. The approach is based on a Knowledge Graph for Ecoinvent data, focused on the explicit and formal representation of LCA indicators and their associated mathematical calculation formulas. The LCA KG is aimed to support an accurate calculation of environmental impacts, thus supporting its certification process by facilitating the documentation of the assessment and its sharing. The schema underlying the KG is based on several ontologies, including the Ecoinvent vocabulary and KPIOnto for representing indicators and related formulas. On top of the model, we propose reasoning-based services based on an algebra system to perform several supporting tasks. Primarily, it enables the automatic calculation of impact indicators from enterprise data sources through algebraic manipulation of formulas. Additional services allow the derivation of alternative calculation expressions for a given indicator and, in case of collaborative enterprises, the calculation of dependencies among indicators across different organizations to ensure consistent comparisons of sustainability assessments.

The remainder of this article is structured as follows: Section 2 discusses the design of the LCA Knowledge Graph, while Section 3 introduces logic-based services providing capabilities useful for formula calculation and comparison of impact results. Finally, Section 4 draws conclusions and outlines future work.

2. Data model

This section aims to discuss the design of the LCA Knowledge Graph, starting from the information stored in the Ecoinvent database and its Data Glossary representing its Linked Data representation. Then, we discuss the KPIOnto ontology for the representation of mathematical indicators and its use for the definition of the integrated schema for the Knowledge Graph.

---

[2] https://ecoinvent.org/

2.1. Background

The Ecoinvent database contains a large number of datasets, organized by multiple categories, such as sector (e.g., Chemicals, Electricity,), geographical area, and activity type (e.g., a transforming activity). The building blocks of the database are the so-called Unit Processes (UPR), also named activities. Each activity represents a single human process that can occur in a supply chain and is characterized by flows, namely elementary exchanges (EEs) and intermediate exchanges (IEs). The first are exchanges from/to the environment and refer to the case in which an activity consumes natural resources or releases emissions. The latter are flows with other human activities and do not involve exchanges with the environment. IEs may output different products or wastes, including a reference product, that is the main driver of the activity. Exchanges have a set of properties, such as the water or carbon content that may serve for analysis e.g., the carbon footprint of the related product. All these information can support LCA with the evaluation of potential impacts to the environment of a specific activity. An interesting piece of information are the mathematical formulas associated with exchanges, properties or parameters. Parameters are specific type of values that can be used to express relationships, such as the efficiency, the correlation of input to variables and the estimation on emissions. Several quantitative measures can be expressed with mathematical relationships, including the amount of exchanges.

For example, let us consider the activity *silicone product production*[3]. Among its EEs, two relate to water emissions respectively to the air (*w_t_a*) and to water (*w_t_w*). Both EEs are associated with a mathematical expression calculating their amounts in the activity (expressed in m^3), as shown below:

- $$amount(w\_t\_a) = \left(\frac{\text{tap\_water\_input}}{1000 * \text{fraction\_TW\_to\_air}}\right) +$$

$$+(\text{water\_cooling\_UNO\_input} * \text{fraction\_CW\_to\_air}) + (\text{water\_well\_in\_ground\_input} * \text{fraction\_PW\_to\_air})$$

- $$amount(w\_t\_w) = \left(\frac{\text{tap\_water\_input}}{1000}\right)(1 - \text{fraction\_TW\_to\_air}) +$$

$$+\text{water\_cooling\_UNO\_input}(1 - \text{fraction\_CW\_to\_air}) + \text{water\_well\_in\_ground\_input}(1 - \text{fraction\_PW\_to\_air})$$

The tap_water_input refers to an IE which exchanges with another activity whose reference product is tap water. The water_cooling_UNO_input is an EE corresponding to the input of cooling water (where UNO stands for 'unspecified natural origin'). The water_well_in_ground_input is another EE referring to the input of well water from the ground. The fraction_TW_to_air, fraction_CW_to_air and fraction_PW_to_air are parameters quantifying the amount of tap water, cooling water and process water emitted to the air, respectively. Another mathematical relationship expresses the equivalence between fraction_TW_to_air and fraction_PW_to_air. Also, fraction_CW_to_air can be calculated as (0.5 * fraction_CW_OT_to_air) + (0.5 * fraction_CW_R_to_air), where fraction_CW_OT_to_air and fraction_CW_R_to_air are the amount of cooling water emitted to the air, respectively once-through system and recirculating system.

Besides being used to characterize how to calculate the amount for a flow, formulas can be expressed also to calculate overall impact indicators, which is essential to conduct LCIA of an activity or the entire supply chain. In particular, Ecoinvent provides documentation for the calculation of LCIA score of activities or individual exchanges, starting from the selection of an impact method. An impact assessment method can include different impact indicators, focusing on the evaluation of single footprints or impact categories. The LCIA score s_i of an activity with reference to a selected impact category i can be computed as: $s_i = \sum_k a_k C_{i,k}$, where a_k is the amount of the flow k which impacts on i, and C is the coefficient. For instance, chosen the impact method EF, it is possible to assess the impact category 'water use' of the *silicone product production* activity by considering all the flows involved in water consumption. Thus, their amounts are multiplied with the corresponding coefficients to obtain the LCIA score which evaluates the environmental impact of the activity on water usage. Although Ecoinvent provides averaged amounts to calculate approximate LCIA scores, a more precise calculation, relying on actual enterprise data, is essential to precisely certify environmental impacts.

---

[3] https://ecoquery.ecoinvent.org/3.10/cutoff/dataset/6368/documentation

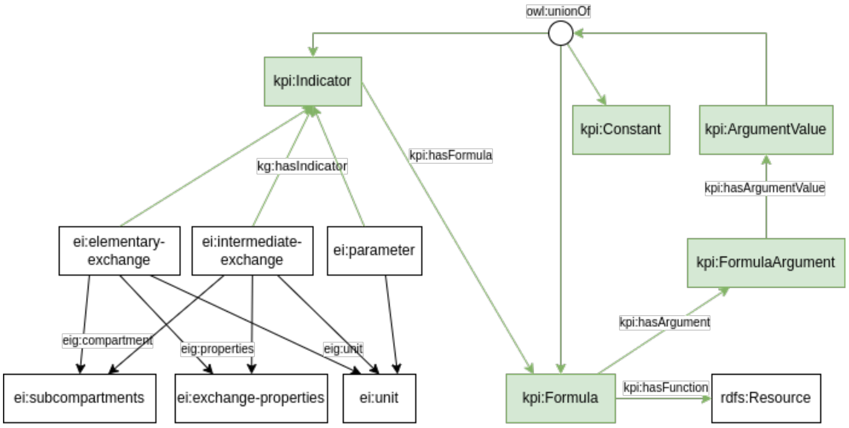


Figure 1: Integrated schema. Prefixes `ei` and `eig` refer to Ecoinvent namespaces, prefix `kpi` refers to the KPIOnto namespace, while the prefix `kg` refers to the custom vocabulary used for integration.

2.2. Ecoinvent vocabulary

The Ecoinvent Data Glossary[4] provides a reference for all metadata used within Ecoinvent inventory datasets, in the JSON-LD format, with the purpose to establish a common understanding of data. Its schema includes classes representing basic concepts such as `elementary-exchange` and `intermediate-exchange` with their description, classification and formula, `exchange-properties` and `parameter` with the `unit` of measurement, and `activity-name`. At the time of the writing, with respect to the information available in the Ecoinvent datasets, the schema is incomplete and data is missing. For example, the relation between an activity and the exchanges is not defined, mathematical formulas are not reported for any flow, and activities' properties are not detailed. Furthermore, the schema assumes formulas are encoded just as a string, namely a `Text` datatype from schema.org.

2.3. KPIOnto

In order to explicitly represent the mathematical expressions to calculate amounts for Ecoinvent flows, we resort on the KPIOnto ontology[5]. The ontology provides classes and properties for describing (Key) Performance Indicators on the Semantic Web. The main class of the ontology is `Indicator`, which describes the quantitative metrics enabling performance monitoring. It is described through a set of properties including a description (`KPIDescription`), a unit of measurement (`unitOfMeasure`), the aggregation function (`aggrType`), its business objective (`hasBusObj`), a set of dimensions of analysis. An indicator can be either atomic or compound, built by combining several lower-level indicators. Dependencies of compound indicators on their building elements are defined by means of algebraic operations, that is a `Formula` capable of expressing the semantics of an indicator compositionally. A formula describes how the indicator is computed and is characterized by the operator (property `hasFunction`) and one or two operands (respectively, for unary and binary operators), which are instances of `FormulaArgument`. In turn, each argument has a value, represented as an instance of `ArgumentValue`. Finally, a value can either be a `Constant`, another indicator or, recursively, a formula. The ontology has been used for various applications, ranging from formally representing a library of KPIs [8], to reasoning on ESG indicators [9], to supporting self-service data analytics [10].

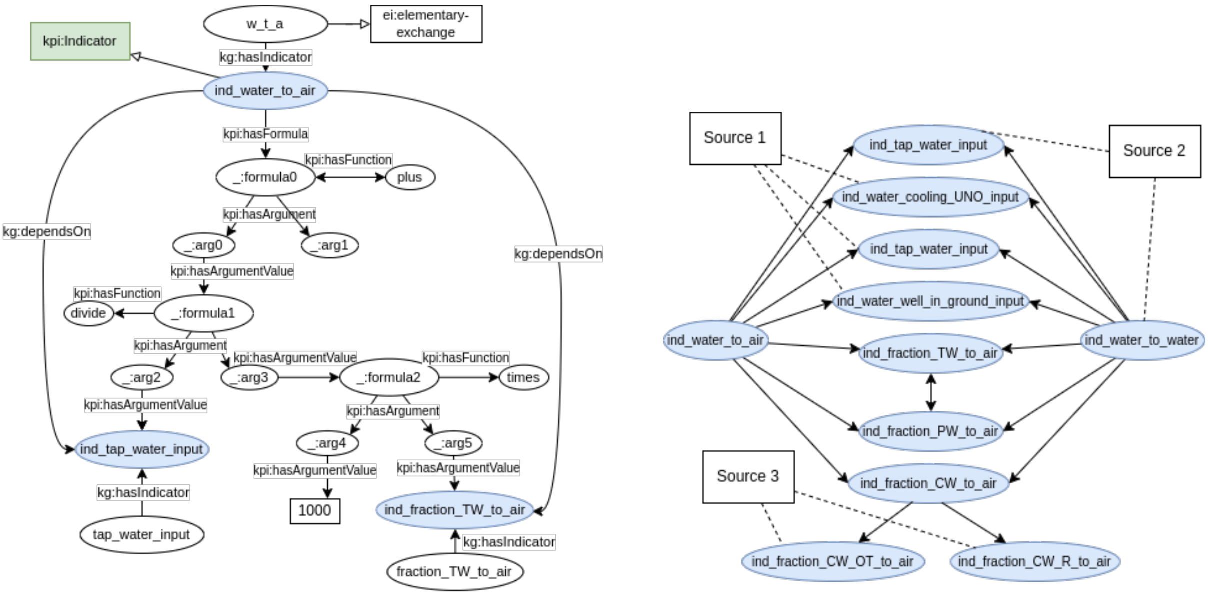
2.4. Integrated schema

The LCA Knowledge Graph has been defined by referring to a schema integrating the Ecoinvent vocabulary and the KPIOnto ontology. As shown in Figure 1, we introduce a property `hasIndicator` from a custom vocabulary (with namespace `kg`) associating each elementary/intermediate exchange and

---

[4] https://glossary.ecoinvent.org/

[5] http://w3id.org/kpionto



(a) (b)

Figure 2: Excerpts of the KG for the *silicone product production* activity: (a) entities related to the water_to_air elementary exchange; (b) `kg:dependsOn` relations among indicators, and mapping to organization data sources.

property with its amounts, represented as an instance of class `Indicator`. In this way, the information related to the calculation of flow amounts is represented using KPIOnto terminology. Furthermore, we introduce a relation `kg:dependsOn` between two indicators, which is defined through a SWRL rule, stating that the former indicator is compound and the latter is one of its components.

In Figure 2a an example of the resulting LCA Knowledge Graph is shown, representing, for lack of space, a subset of the indicators and formulas related to the elementary exchange *water_to_air*, for the *silicone product production* activity. The indicator *ind_water_to_air* is represented as a compound indicator with a formula defined as the application of an operand to a set of operators, which in turn are other indicators, e.g. *ind_tap_water_input* or *ind_fraction_TW_to_air*.

The model is designed to be general and versatile, serving multiple scenarios across different organizations. In practical situations, an enterprise may have data available for only some of the indicators in the model. In such cases, a mapping can be defined to explicitly state that a particular data source contains data related to a specific indicator, following the model proposed in [11]. An example of such mappings is represented in Figure 2b, where a set of indicators are shown together with `kg:dependsOn` relations among them, and linked to the enterprise data sources including corresponding data, e.g. “Source 2” includes data on indicators *ind_water_to_water* and *ind_tap_water_input*.

3. Services for Life-cycle Assessment

The LCA Knowledge Graph can enable an easier documentation of indicator definitions and LCIA results, ultimately making their understanding easier and shareable on the Web using open and a FAIR format. A set of services can be built on top of the LCA Knowledge Graph to support (1) quantitative impact assessments and (2) advanced analysis and comparison of LCIA.

A quantitative calculation of the impact is needed for more precise assessments and to validate the accuracy and consistency of data against regulatory standard, reducing the risk of non-compliance due to data errors. The calculation of indicators relies on an algebra reasoning system, derived from PRESS [12] and implemented in Logic Programming[6]. Reasoning about indicator formulas is mainly based on symbolic and algebraic manipulation according to mathematical axioms (e.g., commutativity,

---

[6] https://github.com/KDMG/PRESS4KPI

associativity and distributivity of binary operators) needed to solve equations. A basic service consists of deriving alternative formulas to calculate a given indicator, by manipulation of the formulas for the indicators defined in the Graph, e.g., by reverting the existing formula *ind_fraction_CW_R_to_air* can be calculated as (*ind_fraction_CW_to_air - ind_fraction_CW_OT_to_air*).

When indicator definitions are mapped to the corresponding data sources at enterprise level, the formula manipulation functionality can be exploited to support the organization in calculating an indicator from the available data. To make an example, let us image an organization needs to calculate the indicator *ind_water_to_air* for the corresponding elementary exchange. Let us suppose the set of indicators available to the enterprise are the seven provided by the data sources in Figure 2b. As such, the formula for the required indicator cannot be directly computed, since there are three indicators which are not provided by any source, namely *ind_fraction_TW_to_air*, *ind_fraction_PW_to_air*, and *ind_fraction_CW_to_air*. However, by considering the indicators' formulas, the algebra system derives that *ind_fraction_CW_to_air* can be calculated by summation of *ind_fraction_CW_OT_to_air* and *ind_fraction_CW_R_to_air*, both in "Source 3". Then, using such calculated indicator and by reverting the formula for *ind_water_to_water*, it is possible to derive *ind_fraction_TW_to_air* and its equivalent indicator *ind_fraction_PW_to_air* from "Source 1", "Source 2". Given that all the needed indicators are either available or a derivation from the available sources has been calculated, the target indicator *ind_water_to_air* can be finally calculated. Conversely, if the enterprise data sources cannot provide the necessary data to compute a target indicator, the service can assist in identifying the missing indicators. This functionality is crucial for organizations to understand their data gaps better and enhance their data collection and management policies.

A suite of more advanced reasoning services can be used for facilitating collaboration among a network of enterprises, enabling them to more easily perform consistent sustainability assessments. This is particularly important for business processes spanning multiple enterprises, in which the overall assessment requires the parties to share relevant indicators in order to calculate the impact indicator at hand. A set of reasoning services are able to determine what indicators are *computable* starting from those already available, and what are the *common dependencies* for the target impact indicators to assess. On its top, a *dependency calculation* service allows to compare the available indicators for each collaborating enterprise, in order to determine the set of common (available or computable) indicators among them, namely the indicators that can be used for making assessments at network level.

4. Conclusion

This article presents an approach based on a formal, shareable representation for LCA data from the Ecoinvent database, which can facilitate the computation and comparison of related indicators and the formal verification of the compliance in evaluating LCIA. By leveraging Knowledge Graphs, our framework supports mapping internal data sources to LCA indicators, thereby aiming to improve the accuracy and reliability of the assessments, while reasoning services can be used for automatic identification of computable indicators and aggregation of basic indicators into compound indicators. This may provide valuable support for collaboration among enterprises, enabling joint sustainability assessments and effective decision-making. Future work will focus on evaluating the approach on real-world cases, expanding reasoning services, integrating additional LCI databases, and exploring applications in various industrial sectors.

Acknowledgments

Cristina Rossetti has received funding from the MUR – DM 118/2023 as part of the project PNRR-NGEU.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] ISO 14040: 2006, SO 14040-Environmental management–life cycle assessment–principles and framework, Standard, International Organization for Standardization, Geneva, CH, 2006.

[2] S. Hellweg, L. Milà i Canals, Emerging approaches, challenges and opportunities in life cycle assessment, Science 344 (2014) 1109–1113.

[3] S. Mohamed, Z. Yingzhong, T. Jinghai, J. Jia, A graph database for life cycle inventory using neo4j, Journal of Cleaner Production 393 (2023).

[4] T. Peng, L. Gao, R. S. Agbozo, Y. Xu, K. Svynarenko, Q. Wu, C. Li, R. Tang, Knowledge graph-based mapping and recommendation to automate life cycle assessment, Advanced Engineering Informatics 62 (2024) 102752.

[5] K. Janowicz, A. A. Krisnadhi, Y. Hu, S. Suh, B. P. Weidema, B. Rivela, J. Tivander, D. E. Meyer, G. Berg-Cross, P. Hitzler, et al., A minimal ontology pattern for life cycle assessment data, in: CEUR Workshop Proceedings, volume 1461, CEUR-WS, 2015.

[6] B. Kuczenski, C. B. Davis, B. Rivela, K. Janowicz, Semantic catalogs for life cycle assessment data, Journal of cleaner production 137 (2016) 1109–1117.

[7] A. Ghose, M. Lissandrini, E. R. Hansen, B. P. Weidema, A core ontology for modeling life cycle sustainability assessment on the semantic web, Journal of Industrial Ecology 26 (2022) 731–747.

[8] C. Diamantini, D. Potena, E. Storti, Sempi: A semantic framework for the collaborative construction and maintenance of a shared dictionary of performance indicators, Future Generation Computer Systems 54 (2016) 352–365.

[9] C. Diamantini, T. Khan, D. Potena, E. Storti, et al., Shared metrics of sustainability: a knowledge graph approach, in: CEUR WORKSHOP PROCEEDINGS, volume 3194, 2022, pp. 244–255.

[10] C. Diamantini, D. Potena, E. Storti, Analytics for citizens: A linked open data model for statistical data exploration, Concurrency and Computation: Practice and Experience 33 (2021) e4186.

[11] C. Diamantini, A. Mele, D. Potena, C. Rossetti, E. Storti, A metadata model for profiling multidimensional sources in data ecosystems, arXiv preprint arXiv:2503.15951 (2025).

[12] L. Sterling, A. Bundy, L. Byrd, R. O'Keefe, B. Silver, Solving symbolic equations with press, Journal of Symbolic Computation 7 (1989) 71–84.