



**HAL**  
open science

## An approach toward a prediction of the presence of asbestos in buildings based on incomplete temporal descriptions of marketed products

Thamer Mecharnia, Lydia Chibout Khelifa, Nathalie Pernelle, Fayçal Hamdi

### ► To cite this version:

Thamer Mecharnia, Lydia Chibout Khelifa, Nathalie Pernelle, Fayçal Hamdi. An approach toward a prediction of the presence of asbestos in buildings based on incomplete temporal descriptions of marketed products. K-CAP '19: Knowledge Capture Conference, 2019, Marina Del Rey, United States. pp.239-242, 10.1145/3360901.3364428 . hal-04471235

**HAL Id: hal-04471235**

**<https://hal.science/hal-04471235>**

Submitted on 26 Feb 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/337271051>

# An Approach Toward a Prediction of the Presence of Asbestos in Buildings Based on Incomplete Temporal Descriptions of Marketed Products

Conference Paper · September 2019

DOI: 10.1145/3360901.3364428

CITATIONS

8

READS

110

4 authors:



**Thamer Mecharnia**

Mines Saint-Etienne

7 PUBLICATIONS 10 CITATIONS

SEE PROFILE



**Lydia Khelifa Ep Chibout**

Centre Scientifique et Technique du Bâtiment

4 PUBLICATIONS 10 CITATIONS

SEE PROFILE



**Nathalie Pernelle**

Université Sorbonne Paris Nord

82 PUBLICATIONS 613 CITATIONS

SEE PROFILE



**Fayçal Hamdi**

Conservatoire National des Arts et Métiers

74 PUBLICATIONS 567 CITATIONS

SEE PROFILE

# An Approach Toward a Prediction of the Presence of Asbestos in Buildings Based on Incomplete Temporal Descriptions of Marketed Products

Thamer Mecharnia

Centre Scientifique et Technique du Bâtiment (CSTB)  
& LRI, Université Paris Sud, Orsay, France  
thamer.mecharnia@cstb.fr

Lydia Chibout Khelifa

Centre Scientifique et Technique du Bâtiment (CSTB),  
Paris, France  
lydia.chibout@cstb.fr

Nathalie Pernelle

LRI, Université Paris Sud, Orsay, France  
nathalie.pernelle@lri.fr

Fayçal Hamdi

CEDRIC, Conservatoire National des Arts et Métiers  
(CNAM), Paris, France  
faycal.hamdi@cnam.fr

## ABSTRACT

Since 1997, the production, import and sale of asbestos<sup>1</sup> have been banned in France. However, there are still millions of tons scattered in factories, buildings, schools and hospitals. In this paper we propose a method for predicting the presence of asbestos products in buildings based on temporal data that describes the probability of the presence of asbestos in marketed products. Our approach was tested on real data based on the 704 asbestos products described in external resources.

## KEYWORDS

Ontology, Uncertain information, Temporal data, Prediction

### ACM Reference Format:

Thamer Mecharnia, Nathalie Pernelle, Lydia Chibout Khelifa, and Fayçal Hamdi. 2019. An Approach Toward a Prediction of the Presence of Asbestos in Buildings Based on Incomplete Temporal Descriptions of Marketed Products. In *Proceedings of K-Cap '19*. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/1122445.1122456>

## 1 INTRODUCTION

For its fireproof qualities, from the pre-war to its ban, France has extensively used asbestos, especially from 1950 till 1970. The harmfulness of asbestos has been known for a long time, but its danger has been identified since the beginning of the 20<sup>th</sup> century. In 1997, the INSERM (a French public research organization dedicated to health) has recalled the different effects of asbestos exposure on health. Since January 1997, the production, import and sale of asbestos have been banned in France. However, there are still millions of tons scattered in factories, buildings, schools and hospitals.

<sup>1</sup>Group of naturally occurring mineral fibres which were used extensively due to their versatility, strength, fire resistance and insulating properties.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*K-Cap '19, 19-22 November 2019, Marina del Rey, California, United States*

© 2019 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM... \$15.00

<https://doi.org/10.1145/1122445.1122456>

As part of the PRDA<sup>2</sup>, the CSTB<sup>3</sup> has been asked to develop an online tool providing assistance in the identification of asbestos-containing materials in buildings which aims to guide the tracking operator in the preparation of his tracking program (ORIGAMI Project). This tool will not replace in any case the identification of materials and products containing asbestos made by professionals on samples collected during an inspection of the building, in accordance with the French NF X 46-020 standard. The problem is that the available building descriptions only contain the classes of used products without giving their references or any other information about them (i.e. valued properties, providers, ...). So, we cannot apply prediction approaches that can learn to classify data based on their RDF description [3, 4].

In this paper, we propose an ontology-based approach that estimates the probability of the existence of asbestos products in a building. This hybrid approach combines statistical and rule-based methods to generate this probability. It relies on the construction year of the building, on reliable but incomplete external resources describing the product references that were in use at this period, and on a set of diagnoses. In section 2, we present the prediction problem and the resources available within CSTB. In section 3, we describe the proposed Asbestos Ontology. Then, in section 4, we present our predictive approach. Finally, section 5 presents the results obtained on a real data set.

## 2 OBJECTIVE

The main objective of the ORIGAMI project is to help the expert predicting the presence of asbestos in products, locations or structures in a given building using the CSTB documentary resources covering the period from 1946 to 1997. The CSTB owns a set of documents describing the buildings built in France, including the following types of documents:

- **Approved type project:** documents containing a set of information describing a building (name, address, type, year of construction, region), and the list of its structures (outside opening, balcony, etc.). For each structure, the document describes all of its locations (door, window, etc.) as well as the product classes used for each location (plaster, glue, etc.).

<sup>2</sup>Asbestos Research and Development Plan

<sup>3</sup>Scientific and Technical Center for Building. <http://www.cstb.fr/en/>

- **Asbestos diagnosis:** documents describing the results of samples taken on a building to detect the presence of asbestos in elements constituting parts of buildings (product types and building elements where the samples have been collected).

In this research project, the challenges are to (1) model an ontology that can describe the buildings and all the knowledge that is required for the asbestos prediction (2) predict the presence of asbestos based on this ontology and on the classes of the used products (e.g. glue) and not on its precise reference (i.e. the name of the glue that has been used).

### 3 ASBESTOS ONTOLOGY

We manually constructed the upper part of the Asbestos Ontology (see Figure 1) based on the CSTB documentation resources, the needs in terms of prediction and by interacting with domain experts. The main concepts of this ontology are:

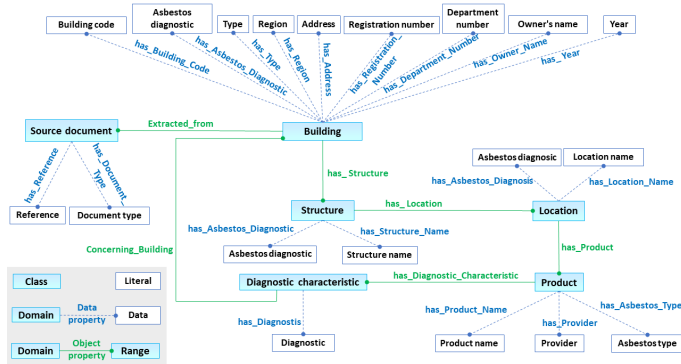


Figure 1: Main Concepts of the Asbestos Ontology

- **Building:** a construction characterized by a code (CSTB internal code that corresponds to a given type of building: school, housing, etc.), a type, the construction year, an address, and its geographic area.
- **Structure:** subspace of the building (e.g. balcony, staircase, roof, floor-ground, etc.).
- **Location:** indicates a basic element that belongs to a building structure (e.g. door, window, wall, etc.).
- **Product:** describes a product that can be used in the composition of locations (e.g. glue, coating, etc.). A product is described by name, provider name and asbestos type.
- **Source Document:** A source document is characterized by its type (approved standard project or asbestos diagnoses) and its reference within the CSTB.
- **Diagnostic Characteristic:** contains the results of the existence of asbestos test.

### 4 PREDICTIVE APPROACH

In this section, we present the predictive approach of asbestos based on our Asbestos Ontology. In CSTB documents, we only know the product class used for a location but the reference of the used product is not given. For example, we know that the used product is a *Coatin* but we do not know which coating (i.e. marketed products)

is exactly used. In what follows, we propose to use a method based on the presence of asbestos in the marketed products to predict the probability of asbestos existence for the products of a building constructed on a given date, then more generally for its locations, its structures, and the building itself.

#### 4.1 Ontology enrichment and population with marketed products

In a first step, we use two external resources provided by ANDEVA<sup>4</sup> and INRS<sup>5</sup> to enrich and populate the Asbestos Ontology with marketed products. ANDEVA publishes on its website<sup>6</sup> a list of 650 products containing asbestos. This list contains the product name and the year from which it is no longer containing asbestos. INRS also publishes a list<sup>7</sup> of 300 products that describes for a set of product names, the time intervals when these products have contained asbestos with a degree of uncertainty.

We enriched the Asbestos Ontology with the ANDEVA and INRS product classes that become sub-classes of the *Product* class. To represent the asbestos characteristics of a product as it is described in each source, as well as the characteristics resulting from the fusion of this information, we added a new concept *Extracted characteristic* that allows us to represent the time interval, the asbestos probability, and the resource of this characteristic (INRS, ANDEVA or their merge) (see Figure 2). We indicated a probability of 1 for products which contain asbestos over the period, 0.5 when asbestos is unknown and 0 otherwise. Since both external resources are equally reliable, we applied a pessimistic approach that considers for each time interval the highest probability of asbestos in the merging process.

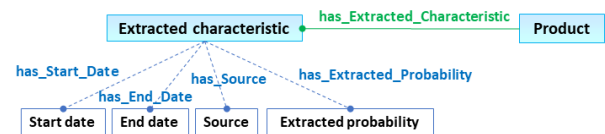


Figure 2: Extracted asbestos features representation

Let's take the example of the *Armazol*. The provenance of the first characteristic of *Armazol* is the INRS source that says that the product contains asbestos (extracted probability  $p_{ext}=1$ ) from 1946 to 1982, then that it potentially contains asbestos until 1997 ( $p_{ext}=0.5$ ). The source ANDEVA says that it contains asbestos from 1946 to 1990 ( $p_{ext}=1$ ), and that there is no asbestos ( $p_{ext}=0$ ) since 1990. The provenance of the third characteristic is the result of the merge of these two characteristics: the product contains asbestos ( $p_{ext}=1$ ) from 1946 to 1990, then it potentially contains asbestos until 1997 ( $p_{ext}=0.5$ ).

#### 4.2 Calculation of the asbestos probability for products

Since the documents do not mention the actually used products during the construction of a building, we assume that we can calculate

<sup>4</sup>National Association for the Defense of Asbestos Victims

<sup>5</sup>National Institute of Research and Security

<sup>6</sup>[http://andeva.free.fr/expositions/gt\\_expos\\_products.htm](http://andeva.free.fr/expositions/gt_expos_products.htm)

<sup>7</sup><http://www.inrs.fr/media.html?refINRS=ED%201475>

the probability of asbestos existence in an unknown used product from the list of marketed products of the same class that exist at the time of this building construction, considering that they are equiprobable. Thus, for each product  $x_k$  belonging to the product class  $C(x_k)$  that is used in a building constructed on a date  $d$ , we calculate a probability of asbestos presence  $p_a(x_k, d)$  by summing the merged extracted probabilities  $p_{ext}$  of the marketed products  $x_j$  belonging to the same class of  $x_k$  and containing asbestos at the same date  $d$ . This sum is then divided by the total number of products of  $C(x_k)$  class that were in use at the same date:

$$p_a(x_k, d) = \frac{\sum_{x_j \in C(x_k)} p_{ext}(x_j, d)}{|x_j|} \quad (1)$$

However, one of the difficulties is that INRS and ANDEVA focus only on products that had asbestos for at least one period. Since we do not have the actual number of the marketed products in a given period of time, we suppose that the total number of the marketed products is largely underestimated and this number may vary depending on the year. We propose to readjust it based on all available sampling diagnoses. Thus, for a given year  $d$ , we compare the proportion of products containing asbestos in the resources regardless of the class product (i.e.  $C(x_k) = Product$ ), with the proportion of products containing asbestos in the diagnostics made for the same year. This allows us to determine what is the  $\alpha$  proportion of the missing asbestos-free marketed products.

$$p_a(x_k, d) \times (1 + \omega) = \frac{\sum_{x_j \in C(x_k)} p_{ext}(x_j, d)}{|x_j| + (\alpha \times |x_j|)} \Rightarrow \alpha = \frac{-\omega}{1 + \omega} \quad (2)$$

In the equation 2,  $\omega$  represents the difference between the calculated probability and the real probability. It results from the comparison of the ratio of products containing asbestos in diagnoses with the ratio of those in the external resources (INRS and ANDEVA) in the same year. It is calculated as follows:

$$\omega = \frac{|p_{diag,a}|}{|p_{diag}|} - \frac{\sum_{x_k \in Product} p_a(x_k, d)}{|x_k|}$$

where  $|p_{diag,a}|$  is the number of products containing asbestos in diagnostics, and  $|p_{diag}|$  is the total number of products in diagnostics.

To model the calculated asbestos probabilities of the unknown product that was used in a building in the Ontology, the concept *Calculated characteristic* has been added (Figure 3). An instance of this concept is described by the calculated probability, the calculation year that corresponds to the year of the building, and the class of the probability. Only two probability classes have been defined : high, and low (coded by 1 and 0 respectively). In fact, given the incompleteness of the data, a probability of 1 or 0 doesn't allow us to predict with certainty that the building contains asbestos or not.

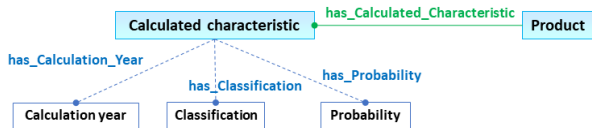


Figure 3: Calculated asbestos features representation

To determine that a product has a high probability to contain asbestos or not, we compare the calculated probability of asbestos presence in this product with a given threshold  $s$ . We use a set of asbestos diagnostics to determine the best threshold, i.e. the threshold that maximizes the accuracy.

### 4.3 Reasoning rules

We have defined SWRL rules that can be exploited by ontology reasoners (1) to propagate high asbestos probabilities that have been associated to unknown class products for a given year to the building elements they belong to (locations, structures, buildings), and (2) to generate missing asbestos probabilities from probabilities that are described for a given year.

To generate the asbestos probability class in a location, a structure or a building, we follow a pessimistic strategy. A rule expresses that a building element is associated to a high probability when at least one of its components has a high probability to contain asbestos.

To generate missing asbestos probabilities for products, we define in the rule pattern R1 (rule pattern defined for each class  $C$ ) that if a product contains asbestos for a given year  $Y_1$ , then it contains asbestos in all previous years  $Y_2 \leq Y_1$ .

**Rule pattern R1:**  $C(?PR1)$ ,  $C(?PR2)$ ,  $hasCalculatedCharacteristic(?PR1, ?C1)$ ,  $hasCalculatedCharacteristic(?PR2, ?C2)$ ,  $hasCalculationYear(?C1, ?Y1)$ ,  $hasCalculationYear(?C2, ?Y2)$ ,  $hasProbability(?C1, 1)$ ,  $lessThanOrEqual(?Y2, ?Y1) \rightarrow hasProbability(?C2, 1)$

Similarly, if a product does not contain asbestos on a date  $Y_1$ , then an instantiated rule pattern allows to infer that it does not contain asbestos for all later years  $Y_2 \geq Y_1$ .

## 5 EXPERIMENTS

The objective of these experiments is to observe how the calculated probability evolves over time and to evaluate our prediction approach on a sample of building descriptions and associated diagnoses.

### 5.1 Evolution of the asbestos probabilities

The aim of this first part of the experiments is to see how the calculated probability of an unknown product evolves depending on its class and on the construction year of the building. To do so, we have first used the diagnoses of 40 products to calculate  $\alpha$  value ( $\alpha \approx 1.0076$ ) which allows us to readjust the number of marketed products. Then, from the merged probability of asbestos existence that has been stored for each marketed product, we have calculated the asbestos probability of an undetermined product for each class and for each year (from 1946 to 1997). The results show that for all classes the calculated asbestos probability remains stable until 1972, then it decreases and reaches 0 in 1997, when asbestos use becomes forbidden. However, the asbestos probability differs from one product class to another. For example, adhesives have lost their asbestos, or have not been marketed, sooner and faster than the other three product classes (see Figure 4).

### 5.2 Evaluation of the predictions on real data

We have evaluated our approach on the available sample of real asbestos diagnoses that describe the tests made on 50 products that belongs to 31 locations, 27 structures and 16 buildings for which

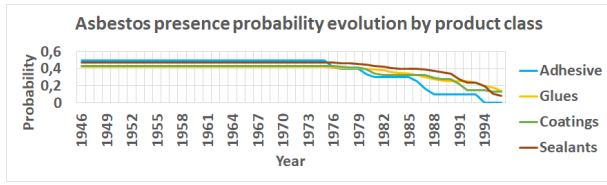


Figure 4: Asbestos probability evolution by product class

Table 1: Confusion matrix for products, locations, structures and buildings

Products		Locations		Structures		Buildings	
TP=34	FP=9	TP=19	FP=6	TP=17	FP=6	TP=15	FP=0
FN=0	TN=7	FN=0	TN=6	FN=0	TN=4	FN=0	TN=1

Table 2: Recall, Precision, F-measure and Accuracy for all types of building component ( $s=0.25$ )

	Recall	Precision	F-measure	Accuracy
Products	1	0.79	0.88	0.82
Locations	1	0.76	0.86	0.81
Structures	1	0.77	0.87	0.78
Buildings	1	1	1	1

the construction year varies from 1949 to 1996. Asbestos has been detected by track-operators in 34 products while 16 of them are asbestos-free, in 19 locations while 12 do not have asbestos, in 17 structures while 10 are asbestos-free, and in 15 buildings among 16.

We first generated the calculated asbestos probability for the products. Then, we applied a leave-one-out method on this set of products to learn the threshold that maximizes the accuracy. We found that the best threshold is the same for all possible subsets of 49 products ( $s = 0.25$ ). Finally, we inferred the products, locations, structures and buildings that have a high probability to contain asbestos. Table 1 represents the obtained confusion matrix for the products, their locations, their structures and the buildings. Table 2 shows that we can correctly classify more than 80% of the products using the chosen threshold (accuracy of 0.82). Furthermore, our pessimistic approach allows us to obtain an accuracy of 0.81 (resp. 0.78) for the locations (resp. the structures) and an accuracy of 1 for the buildings. It means that all the buildings that have a high asbestos probability have at least a component that contains asbestos while all the buildings with a low probability are asbestos-free.

## 6 RELATED WORK

Recent methods of supervised learning attempt to establish predictions using graph data (see [4] for a survey). The difficulties are to know how to represent the structure of the graph, or statistical elements of this structure (e.g. degrees, clustering coefficients) in order to make it exploitable by learning methods[4]. Inductive logic programming and rule discovery approaches [1, 3, 6, 7] can detect rules such as "If a person was born in a country, they probably speak the language of that country". These rules can then be used to predict new information. In this work, we cannot apply such approaches since the RDF description of the product is not available. So, we aim to predict the presence of asbestos in an unknown product

thanks to the properties of all the marketed products belonging to the same class that were in use when the building has been constructed. The rules that allow to propagate the high probabilities to building elements are declared and follow a pessimistic approach which reproduces the reasoning of a building expert in this risk analysis. The available diagnoses are uniquely exploited to learn the threshold used in the classification process of the unknown product and to readjust the number of asbestos-free products that is underestimated due to the incompleteness of the external resources.

The notion of completeness of a knowledge graph does not always make sense according to the properties and classes that one considers, and without a reference graph to compare with [8]. To learn in such a context, some complementary hypotheses have been introduced [2, 3], such as the partial completeness of the properties. Other approaches aim to measure the incompleteness of the data, based on an ideal computed schema [5, 10] or an ideal distribution [9]. In our approach, we evaluate the number of missing asbestos-free marketed products, and therefore the number of specific missing class instances, using the small number of available diagnostics.

## 7 CONCLUSION

This paper presented an approach that combines statistical and rule-based methods to predict the presence of asbestos in a building. We first defined an ontology that models building characteristics and diagnostics. This ontology is enriched by probabilistic temporal data which describes the presence of asbestos in marketed products. We then proposed a pessimistic method to calculate asbestos probabilities based on these uncertain and incomplete data. Experiments performed on real datasets show that our approach achieves good results. They also show that asbestos probabilities changes according to the products classes and the construction year of the building. In future work, we plan to use the calculated probability for the products, the building description, and a larger set of diagnoses to discover the characteristics of buildings, structures, and locations that can also affect the presence of asbestos.

## REFERENCES

- [1] Claudia d'Amato, Andrea G. B. Tettamanzi, and Duc Minh Tran. Evolutionary discovery of multi-relational association rules from ontological knowledge bases. In *EKAW, Bologna, Italy, November 19-23*, pages 113–128, 2016.
- [2] Luis Galárraga, Simon Razniewski, Antoine Amarilli, and Fabian M. Suchanek. Predicting completeness in knowledge bases. *CoRR*, abs/1612.05786, 2016.
- [3] Luis Galárraga, Christina Teflioudi, Katja Hose, and Fabian M. Suchanek. Fast rule mining in ontological knowledge bases with AMIE+. *VLDB Journal*, 24(6):707–730, 2015.
- [4] William L. Hamilton, Rex Ying, and Jure Leskovec. Representation learning on graphs: Methods and applications. *IEEE Data Eng. Bull.*, 40(3):52–74, 2017.
- [5] Subhi Issa, Pierre-Henri Paris, and Fayçal Hamdi. Assessing the completeness evolution of dbpedia: A case study. In *Advances in Conceptual Modeling - ER 2017 Workshops, Spain, November 6-9*, pages 238–247, 2017.
- [6] Nada Lavrac. Inductive logic programming. In *WLP*, pages 146–160. Institut für Informatik der Universität Zürich, 1994.
- [7] Victoria Nebot and Rafael Berlanga Llavori. Finding association rules in semantic web data. *Knowl.-Based Syst.*, 25(1):51–62, 2012.
- [8] Simon Razniewski, Fabian M. Suchanek, and Werner Nutt. But what do we actually know? In *Workshop AKBC@NAACL-HLT, CA, USA*, pages 40–44, 2016.
- [9] Arnaud Soulet, Arnaud Giacometti, Béatrice Bouchou Markhoff, and Fabian M. Suchanek. Representativeness of Knowledge Bases with the Generalized Benford's Law. In *ISWC, Monterey, CA, United States, October 2018*.
- [10] Thomas Pellissier Tanon, Daria Stepanova, Simon Razniewski, Paramita Mirza, and Gerhard Weikum. Completeness-aware rule learning from knowledge graphs. In *IJCAI*, pages 5339–5343. ijcai.org, 2018.