Research on The Evolution of MDE Terminologies based-on Data Mining
One-page desription

Weixing Zhang

September 2021

1 Description

Model-driven engineering (MDE) is a software development methodology that focuses on creating and exploiting domain models, which are conceptual models of all the topics related to a specific problem[1]. In the past several decades, the terminologies in MDE, such as model, modeling language, etc., are constantly emerging and being stereotyped. However, there is a lack of intuitive and comprehensive description of the evolution process of the terminologies. In the following text, this thesis uses "evolution process" to describe the process of MDE terminologies from emergence to being stereotyped, and then to wide use.

Wikipedia is often one of the choices for people who want to learn about the evolution history of a specific terminology because it usually defines the terminology and provides information about its emerging time (or formally defined time). Sometimes it will also introduce the evolution process of the terminology and the reasons behind it. However, there are still deficiencies. For example, Wikipedia is usually unable to introduce terminologies formed a short time ago (such as blended modeling), or lacks a description of some terminologies' evolution (such as domain specific modeling). Moreover, the textual introduction on the evolution history of terminologies on Wikipedia is not intuitive, and it is not simply available when you want to see the relationship between the evolution of many terminologies in MDE.

Google's Books NGram Viewer seems to partially answer these questions, i.e. it tracks the changes of a terminology in the form of a graph (as shown in the Figure 1). However, such a graph doesn't tell too much, e.g. do different terminologies influence each other in the process of evolution.

To make up for the above shortcomings, this thesis aims to use web crawler technology to mine the terminologies in MDE from the internet, and then analyze the data, and make statistics on aspects such as their emerging timeline and the nature of the web pages (i.e. academic or industry?) that mention them, so as to present a more complete view on the evolution of terminologies.

2 Potential RQs

- How long has the terminology in MDE been widely mentioned or applied after it was first mentioned?

- What is the evolution process of terminologies in the MDE?

- How do different terminologies in MDE affect each other's evolution?

3 Methodology

The methodology will be multi-step:

---

[1] https://en.wikipedia.org/wiki/Model-driven_engineering

1996   1998   2000   2002   2004   2006   2008   2010   2012   2014
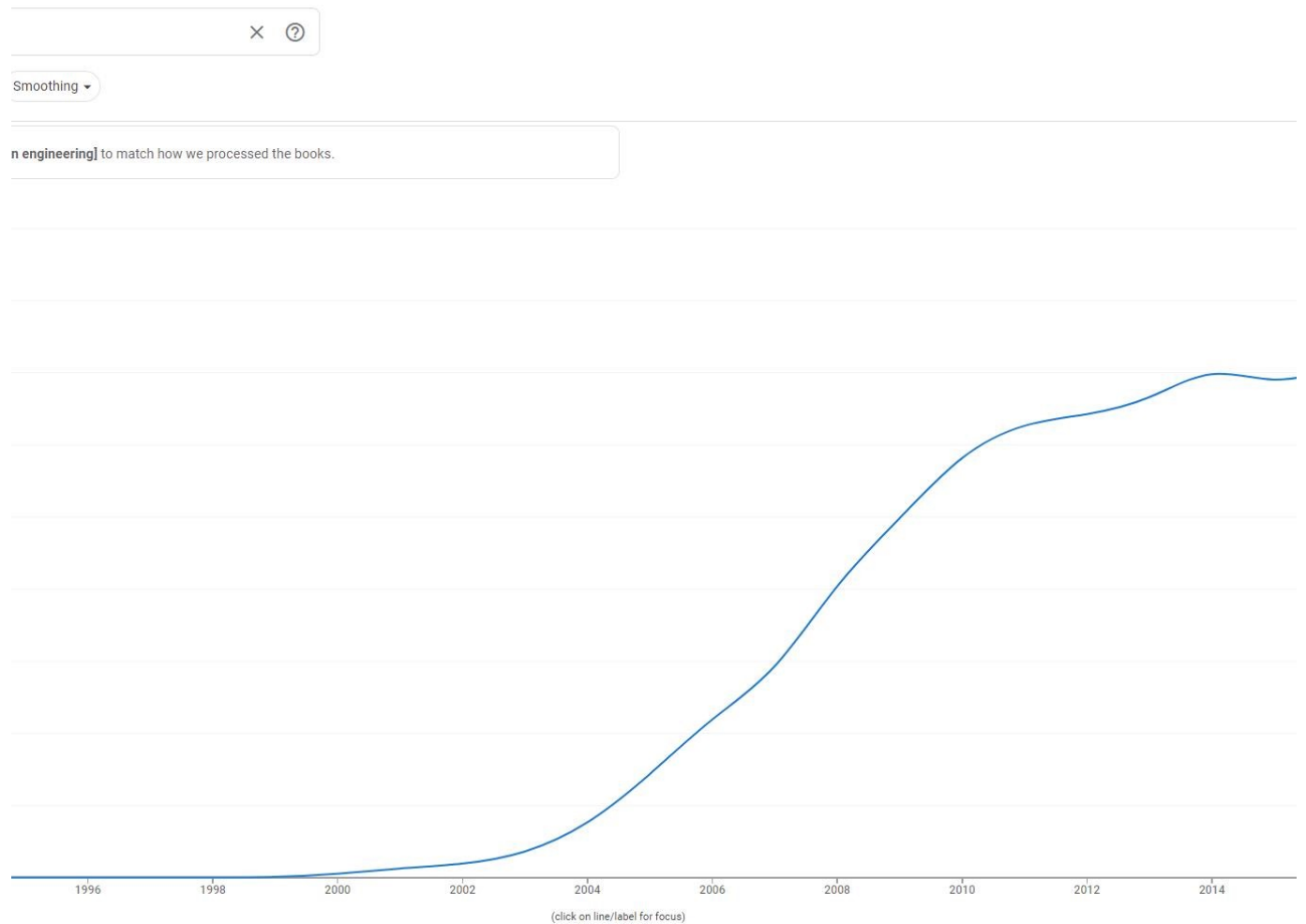
(click on line/label for focus)

Figure 1: Example of Books Ngram Viewer

- First, request a expert group to provide a list of the most important terminologies in MDE, such as "model", "MDE", "DSML", and so on.

- Then, implement a web crawler to capture inforamtion about the existance of terminologies in MDE from the internet. The information includes but is not limited to: the time of existence, the nature of the web pages (academia or non-academic), etc.

- Third step, analyze the collected data. Specifically, apply statistics and visualization of the data, and describe the data combined with manual analysis, so as to describe the evolution process of terminologies in MDE.

4 Required technologies

- Web crawler design and implementation (programming)

- Statistics

Target group

DV, D and IT