



Universiteit
Leiden
The Netherlands

Mining semi-structured data, theoretical and experimental aspects of pattern evaluation

Graaf, E.H. de

Citation

Graaf, E. H. de. (2008, October 29). *Mining semi-structured data, theoretical and experimental aspects of pattern evaluation*. Leiden Institute of Advanced Computer Science, Faculty of Science, Leiden University. Retrieved from <https://hdl.handle.net/1887/13207>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/13207>

Note: To cite this publication please use the final published version (if applicable).

Stellingen (Propositions)

by Edgar de Graaf, author of

Mining Semi-Structured Data

Theoretical and Experimental Aspects of Pattern Evaluation

1. Not only the number of times a pattern occurs, but also the way it occurs makes the pattern interesting. [this thesis]
2. Domain specific knowledge can speed-up highly optimized data mining algorithms. [this thesis]
3. By visualization of co-occurrence a user can quickly see substructures that occur many times in combination. [this thesis]
4. When we mine one dataset using different techniques suited for different structures where transaction arrival time is of importance, then we need to have access to the relative order of items and transactions. [this thesis]
5. Patterns with many occurrences are often obvious and / or well known. The thresholds (minimal occurrence or others) need to be designed and set such that some low occurring interesting patterns are also discovered within reasonable time.
6. If a high ratio of the patterns is interesting to the end-user then a visualization of understandable patterns should be the end-product of any data mining application.
7. In Data Mining research the focus should be on finding interesting patterns in a reasonable time. Finding patterns in optimal time should only be a secondary goal.
8. Presenting an empirical runtime performance comparison with other algorithms is useful, but not always possible.
9. Interdisciplinary research often yields nice results. However, setting up a real and effective co-operation is difficult.