University of Zurich[UZH]

Department of Informatics

University of Zurich
Department of Informatics
Binzmühlestr. 14
CH-8050 Zürich
Phone. +41 44 635 43 11
Fax +41 44 635 68 09
www.ifi.uzh.ch/dbtg

UZH, Dept. of Informatics, Binzmühlestr. 14, CH-8050 Zurich

First_name Last_name

Switzerland

Prof. Dr. Michael Böhlen
Professor
Phone +41 44 635 43 33
Fax +41 44 635 68 09
boehlen@ifi.uzh.ch

Zürich, 25. September 2012

SQL Implementation of the Centroid Decomposition Method

Informatik-Vertiefung (3 ECTS):

Work overview:

The aim of this Vertiefung is to investigate and implement the Centroid Decomposition method using SQL queries. The Centroid Decomposition is a matrix decomposition technique that decomposes an input matrix $\mathbf{V}$ of data into a sum of the product of the loading vectors denoted as L and the centroid vectors denoted as C^T. Formally, a matrix $\mathbf{V} = [V_1|V_2|\ldots|V_n] \in \mathcal{R}^{m\times n}$ can be decomposed as follows: $\mathbf{V} = \sum_{i=1}^{d} L_i C_i^T$ where:

- L: Loading vectors
- C^T: Centroid vectors
- d: number of centroid vectors to compute

A description of the algorithm together with the Java implementation of the Centroid Decomposition will be provided as an input for this work. An empirical comparison between the SQL implementation and the Java implementation should be provided as final output for this Vertiefung.

Work tasks:

1. Understand and implement the Centroid Decomposition algorithm using PL/SQL
2. Evaluate the scalability of the implementation on an Oracle server: horatio.ifi.uzh.ch. The test datasets are already loaded on the server.
3. Empirical comparison of the running time between the SQL implementation and the Java implementation
4. Report of 5-10 pages

5. Oral exam (approx. 25 min)

Literature:

1. Chu, M.T., and Funderlic, R.E., : *The Centroid Decomposition: Relationships Between Discrete Variational Decompositions and SVDs*, in SIAM J. Matrix Analysis and Applications, 2002
2. Navas, M., and Ordonez, C., *Efficient computation of PCA with SVD in SQL*, in DMMT, 2009

Task assignment and supervisor:

- Mourad Khayati

Starting date of thesis: TBD

Ending date of thesis: TBD

University of Zurich
Department of Informatics

Prof. Dr. Michael Böhlen
Professor