Abstract of **"Works of Hans-Peter Kriegel"** presentation at the research group seminar
(original presentation in Russian)

Artem Sevastopolsky, 2015

Presentation *"Works of Hans-Peter Kriegel"* was prepared at the seminar of A.G. D'yakonov research group.

Hans-Peter Kriegel is German computer scientist working in Ludwig-Maximilians-Universitat in Munich, Germany. He specializes in cluster analysis, anomaly detection, methods for spatial data analysis. In 2015 Hans-Peter Kriegel has received prestigious ACM SIGKDD Innovation Award for his contributions such as DBSCAN, OPTICS, LOF, R*-tree algorithms.

The presentation covers DBSCAN (Density Based Spatial Clustering of Applications with Noise) algorithm in detail (article **"A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise"** by Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, 1996). Main definitions and theorems presented in the original article are given in the presentation. Also a few experiments on different data sets were presented in order to show how algorithm performs on data with clusters of various shapes.

Another topic covered by presentation is approach to online news categorization and discovering global and local bursts in a stream of news proposed in **"Discovering Global and Local Bursts in a Stream of News"** article by Max Zimmerman, Irene Ntoutsi, Zaigham Faraz Siddiqui, Myra Spiliopoulou, Hans-Peter Kriegel, 2012. The article contains description of approach to news categorization and architecture of storage for the latest news articles. The storage consists of varying number of big containers for wide news topics, each of which contains small containers for local topics. Each big container also contains a place for news documents that show novelty. When size of these containers is exceeded, it is considered as a burst in a stream of news. The architecture of the storage and all stages of authors' solution were presented, including necessary algorithms such as Fuzzy C-means clustering algorithm and TF-IDF statistic.