**RESEARCH PAPER ANALYSIS**

**By: Kaustubh Chakravarty (Enrollment No: 16103080)**

The following research papers were thoroughly studied and analyzed. Mentioned below are a particular paper’s salient features and a summary of its contents:

1. **Tag-LDA for Scalable Real-Time Tag Recommendation**
   1. **Author(s):** Xiance Si, Maosong Su
   2. **Year:** 2009
   3. **Dataset:** Real world Chinese blogs, and particularly restricted to those blogs only.
   4. **Proposed Algorithm/Model:** A probabilistic model named tagged Latent Dirichlet Allocation (tag-LDA). Main role of this model is to link tags to the topics of a document in an attempt to capture a tag’s semantic meaning using a probability factor and utilizing Gibbs Sampling.
   5. **Method:** When the system starts, it reads all document independent tag-LDA model parameters into the memory and finds the probabilities. Among the recommendations, the top 1000 tags are returned as the result.
   6. **Result:** Tags that occurred less than 10 times were not reported. Improved performance as compared to other available methods (available at the time).
   7. **Critical Analysis:** Unable to discern the semantic meaning of a tag. Also, performance (and precision) drops exponentially as size of content increases.
2. **Real Time Automatic Tag Recommendation**
   1. **Author(s):** Yang Song, Ziming Zhuang, Huajing Li, Qian Kun Zhao, Jia Li, Wiang-Chien Lee, C. Lee Giles
   2. **Year:** 2011
   3. **Dataset:** Primarily considered two real world datasets- **CiteULike** (A website for scientific documentation) and del.icio.us (for webpages).
   4. **Proposed Algorithm/Model:** Consists of a two-step process. First step is to divide the contents into a bipartite graph. Second step is to impose upon it the Two-way Poisson mixture model in order to evaluate the best nodes.
   5. **Method:** This model considers the relationship between documents, words and tags represented within the model via two bipartite graphs to form an overarching weight matrix. After this step, similarity between the test document and the training documents is considered, followed by, segregation by the Two-way Poisson distributed method- to simultaneously cluster words and classify documents.
   6. **Result:** Exact details of the result are omitted in the paper. However, they are evaluated using the following evaluation metrics:
      * Top-k accuracy**:** Percentage of documents correctly annotated to by at least one of the top kth returned tags.
      * Exact-k accuracy: Percentage of documents correctly annotated by the kth recommended tag.
      * Tag-recall: Percentage of correctly recommended tags among all tags annotated by the users.
      * Tag-precision: Percentage of correctly recommended tags among all tags recommended by the algorithm.
   7. **Critical Analysis:** Performance can be improved by changing the individual multiplying step to repeated addition step in the probability factor finding phase. Also, factors of learning are absent.
3. **Learning in Efficient Tag Recommendation:**
   1. **Author(s):** Masek Lipczak, Evangelos Milios
   2. **Year:** 2013
   3. **Dataset:** Considers dataset of three collaborative tagging systems mentioned below:
      * BibSonomy: A repository of webpage bookmarks and references to scientific publications.
      * Delicious: A popular social bookmarking site.
      * Stack Overflow: A popular “Question and Answer” forum for programmers.
   4. **Proposed Algorithm/Model:** The model is made up of 5 basic recommenders- Title recommender, Title-to-tag recommender, Tag-to-tag recommender, Resource profile recommender and User Profile recommender. How they function coherently is described in the section below.
   5. **Method:** Tags extracted from title are taken as input and run a spreading activation algorithm using title-to-tag or tag-to-tag co-occurrence graphs. Such a result is produced by each recommender system (mentioned above in the Proposed Model heading)- known as a tag recommendation set (a set of proposed tags with assigned scores). This is followed by the merging stage which selects 2 tag recommendation sets and linearly re-scores the tags via a given merge coefficient.
   6. **Result:**
      * Enhances generality, adaptability and efficiency of the model via automatic parameter tuning.
      * The model has the system architecture of a text indexing machine with an added cache that makes real-time tagging an efficient process.
   7. **Critical Analysis:** Personalized tagging not yet realized. Performance of model can also be further be improved using a database system.