



Synonym Based Tag Cloud Generation

Charibeth Cheng*, Therese Angustia, Mark Henry Ching,
Cara Andrea Cristobal, Germaine Marie Gabuyo
¹ De La Salle University
*chari.cheng@delasalle.ph

Abstract: A tag cloud is a text-based visual representation of a set of tags which usually depicts the tag's importance in a given text. The presentation and layout of tags can be controlled so that features such as the size, font and color can be used to give some measure of importance of a given tag. Words that are used frequently will be displayed with an increased font size; while tags may appear in uniform or varying colors for aesthetics purposes or otherwise. The purpose of a tag cloud is to allow one to see, at a glance, the content of a document. Unfortunately, existing tag cloud generators produce clouds with tags that do not contribute in identifying the general content of a given document. These generators base the tags its frequency in the document. Thus, there may be tags, which are inflections of the same word, thereby populating the cloud with the same rootword. Furthermore, there may be frequently occurring non-stopwords, but are nevertheless non-discrimating (such as big, good, fine, etc.) SynTag is a tag cloud generator that uses syntactic and relational information of the words in the document in determining the relevant words/phrases in a document. With this, the generated tag cloud only shows tags which give focus on the content rather than illustrating the frequently occurring words. This system addresses the problems of existing term-based and frequencybased tag cloud generators.

Key Words: Tag Cloud, Synonym-based Tags, Word Relations