Semantic Facets
An in-depth Analysis of a Semantic Image Retrieval System

Jonathon S. Hare[1], Paul H. Lewis[1], Peter G. B. Enser[2] & Christine J. Sandom[2]

(1) School of Electronics and Computer Science, University of Southampton

(2) School of Computer, Mathematical and Information Sciences, University of Brighton

THE SEMANTIC FACETS OF AN IMAGE

We have developed a new **faceted model** of semantic content which is a significant advance upon those **conceptual indexing models** previously encountered in the image retrieval literature which classify image semantic content broadly into **generic**, **specific** and **abstract** concepts. Central to our formulation is a combination of **object**, **spatial**, **temporal** and **activity/event** facets. To these are added **abstract** and **related concept** facets, together with **context** and **topic** facets, which capture the highest level, global semantic content of the image.

Generic Object Instance — Bridge
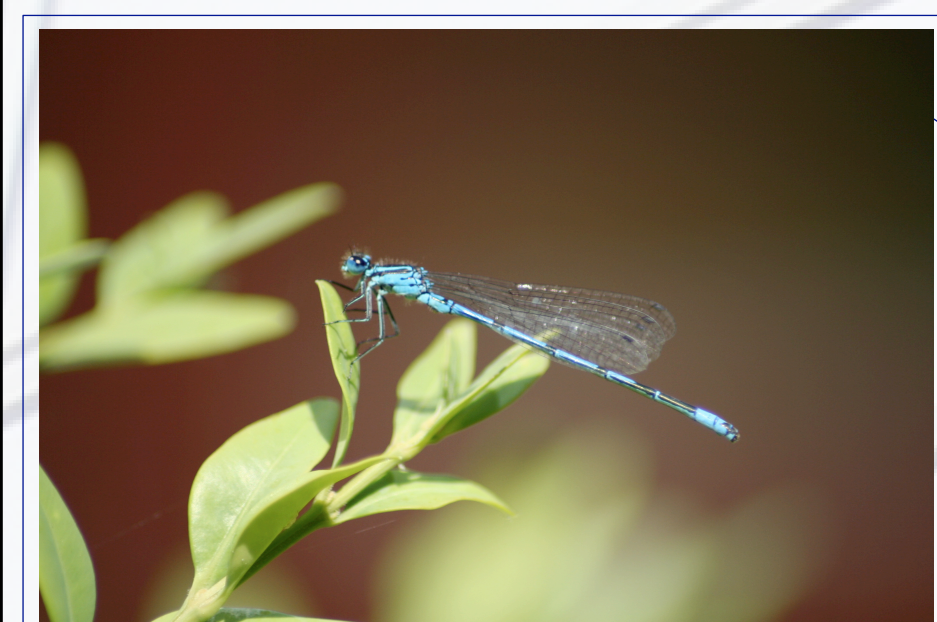Generic Object Class Hierarchy — Red Bridge
Specific Named Object Class — Suspension Bridge
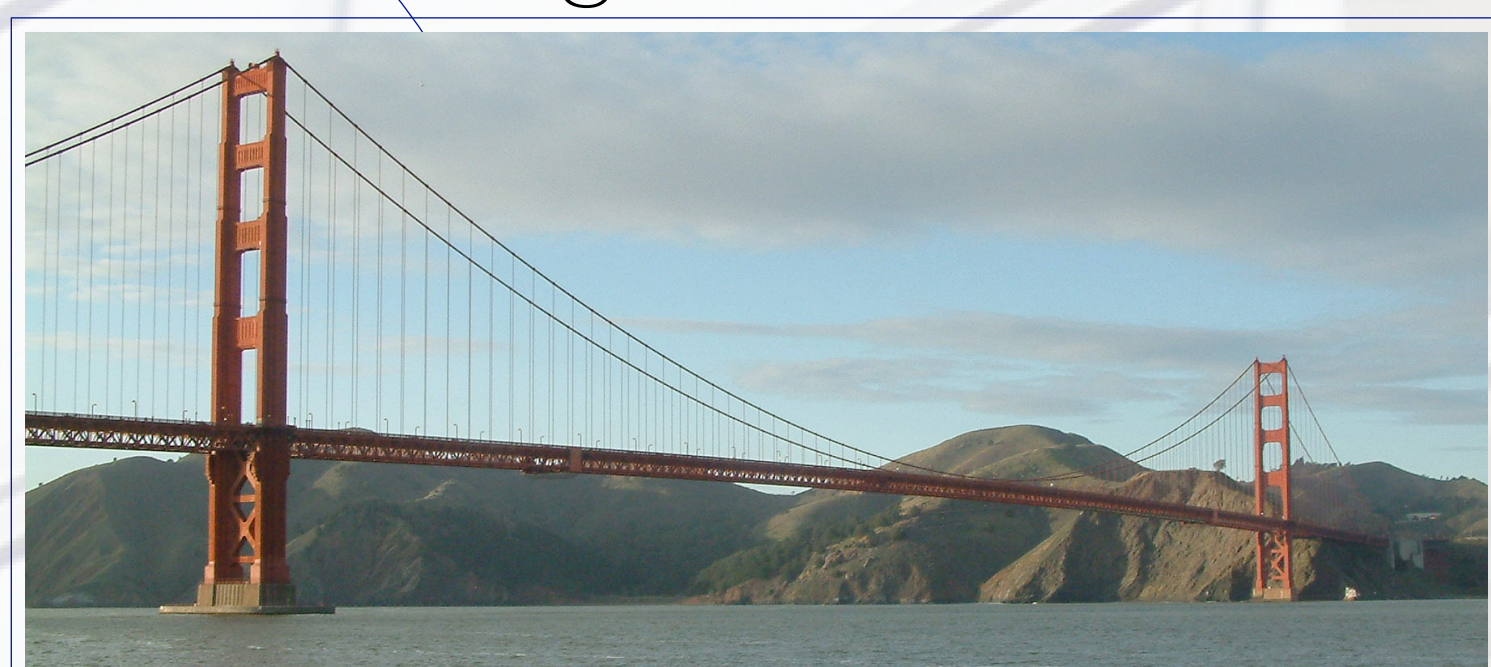Specific Named Object Instance — Golden Gate Bridge
Generic Time — Summer, Morning
Specific Time — 17 June 2006 @ 1046AM
Generic Activity — Standing, Posing
Generic Event — Ceremony
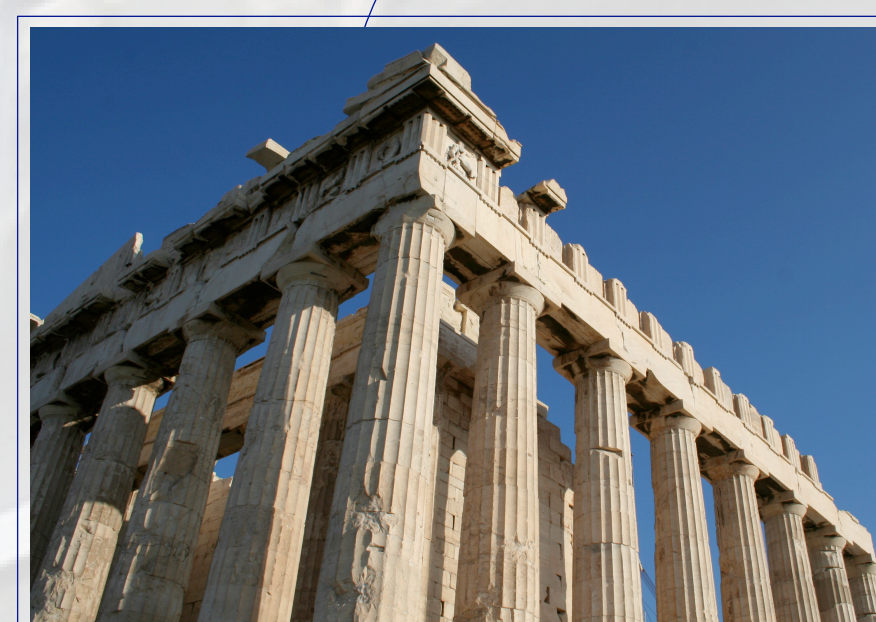Specific Named Event Class — Graduation Ceremony
Specific Event Instance — Comp. Sci. Graduation Ceremony 2006
Specific Location — Acropolis, Athens
Abstract Concept — Ancient Greece
Generic Location — Europe, Greece
Topic — Incarceration
Related Concept — Penal Justice System

SEMANTIC RETRIEVAL

Our **Semantic Space** approach to multimedia retrieval is based on a generalisation of an information retrieval technique called **Cross-Language Latent Semantic Indexing** (CL-LSI).

Conceptually, a Semantic Space is simply a large, **multidimensional vector space** in which images and terms (both keywords and visual-terms describing the images' content) representing the images, are placed. The placement of these images and terms is such that the images are placed 'near' to the terms that describe them. Standard **cosine techniques** can be used to assess and rank similar items in the space.


THE V&A DATASET

The Victoria & Albert collection is an eclectic collection of photographed images of a large variety of subjects. There are over **17000** images, with a keyword-vocabulary of over **12000** terms. Each image has between **0** and **136** keywords (**average 9.2**) associated with it, describing some of the subject meta-data.
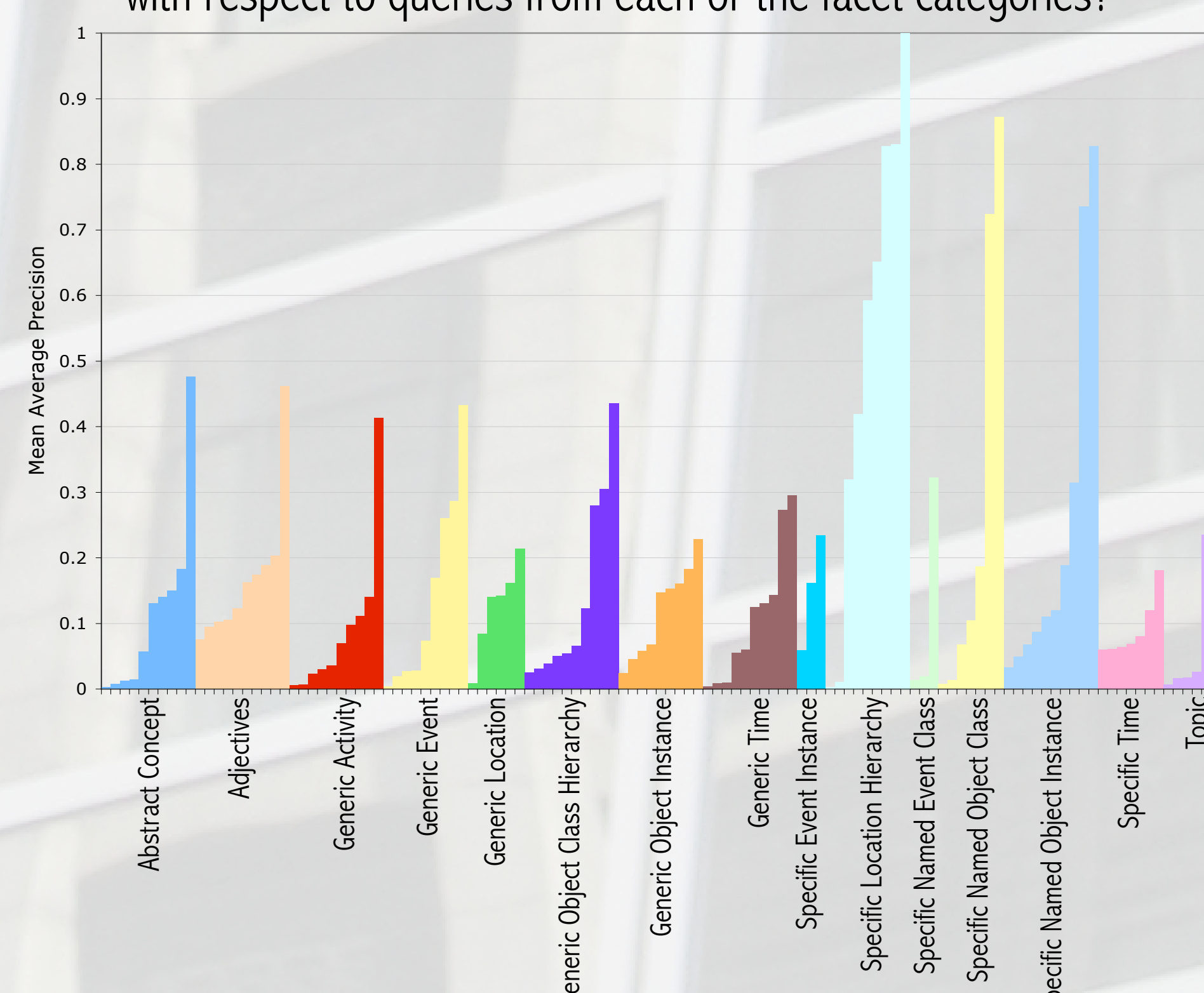

EXPERIMENTS & EVALUATION

Using the **V&A dataset** together with our **faceted model** and **semantic retrieval system**, we set out to investigate three main questions;

Firstly, how is our faceted model reflected within the V&A keywording scheme?
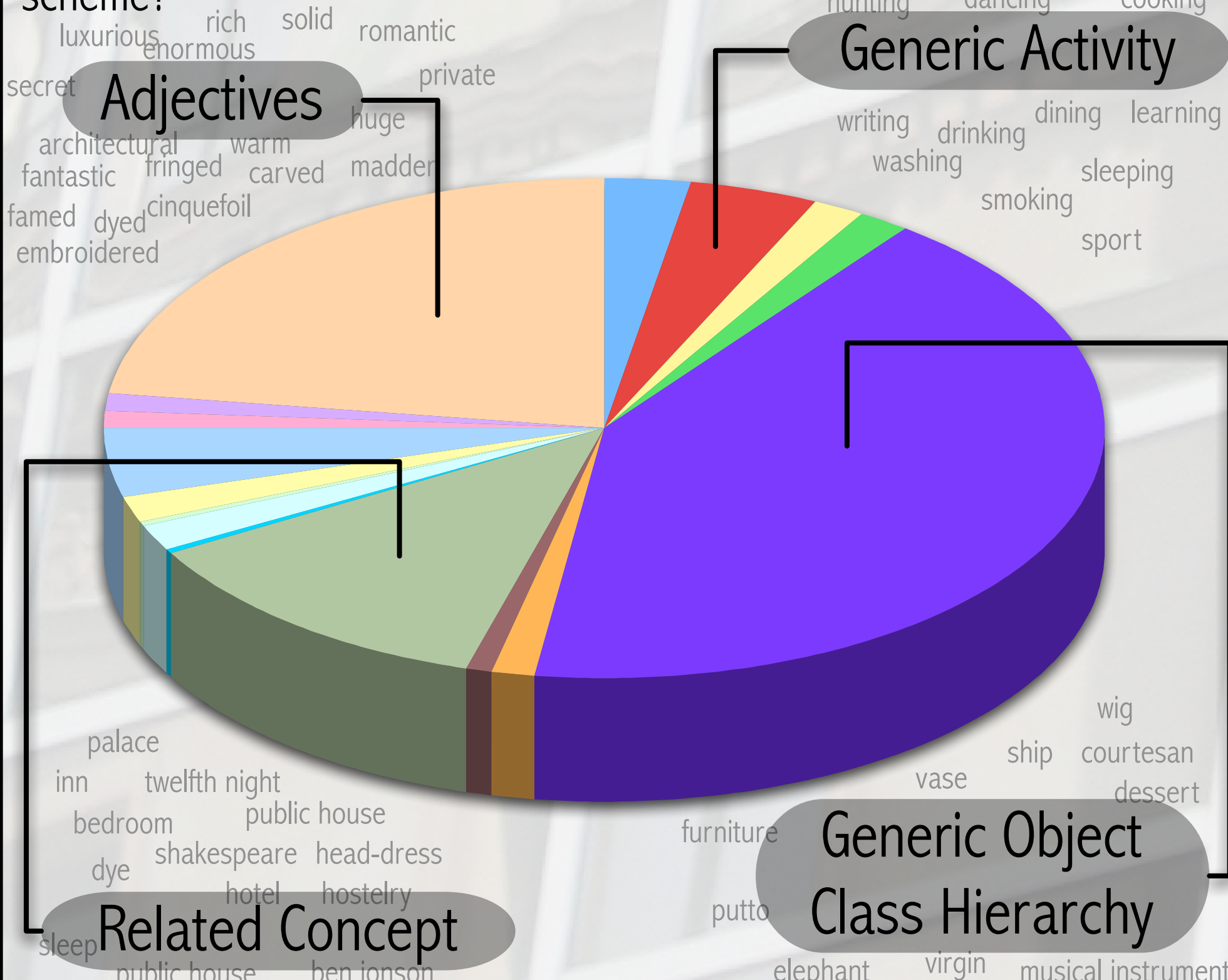
Secondly, how well does the semantic space retrieval system perform with respect to queries from each of the facet categories?

Finally, how does the amount of training data for a particular keyword affect retrieval performance when using that term as a query?