# Generating Summaries and Visualization for Large Collections of Geo-Referenced Photographs


Alexandar Jaffe,
Mor Naaman
Yahoo! Research Berkeley
Berkeley, CA, USA
ajaffe@cs.washington.edu,
mor@yahoo-inc.com

Tamir Tassa
Department of Mathematics
and Computer Science
The Open University of Israel
Ra'anana, Israel
tamirta@openu.ac.il

Marc Davis
Yahoo! Inc.
Sunnyvale, CA, USA
marcd@yahoo-inc.com



## ABSTRACT

We describe a framework for automatically selecting a summary set of photos from a large collection of geo-referenced photographs. Such large collections are inherently difficult to browse, and become excessively so as they grow in size, making summaries an important tool in rendering these collections accessible. Our summary algorithm is based on spatial patterns in photo sets, as well as textual-topical patterns and user (photographer) identity cues. The algorithm can be expanded to support social, temporal, and other factors. The summary can thus be biased by the content of the query, the user making the query, and the context in which the query is made.

A modified version of our summarization algorithm serves as a basis for a new map-based visualization of large collections of geo-referenced photos, called Tag Maps. Tag Maps visualize the data by placing highly representative textual tags on relevant map locations in the viewed region, effectively providing a sense of the important concepts embodied in the collection.

An initial evaluation of our implementation on a set of geo-referenced photos shows that our algorithm and visualization perform well, producing summaries and views that are highly rated by users.


## 1. INTRODUCTION

With the popularization of digital photography, people are now capturing and storing far more photographs than ever before. Indeed, we are moving towards Susan Sontag's 1977 vision of a world where "everything exists to end up in a photograph" [18]. As a result, billions of images, many of which are on the Web, constitute a growing record of our culture and shared experience. Viewing and interacting with such collections has a broad social and practical importance. However, these collections are inherently difficult to navigate, due to their size and the inability of computers to understand the content of the photographs. The prospects of 'making sense' of these photo collections has become largely infeasible.

Some steps forward have been made through geo-referencing of digital photographs, whereby photos are connected to metadata describing the geographic location in which they were taken [12, 19]. Capture devices such as camera-phones and GPS-enabled cameras can automatically associate geographic data with images[1] and will significantly increase the number of geo-referenced photos available online. Already, an increasing number of photographs on the Web are associated with GPS coordinates. Such geo-referenced photos can be categorized geographically or displayed on a digital map, providing a rich spatial context in which to view subsets of a collection. Yet even here, we run into the problem of being able to filter, sort and summarize the data. The viewable space inevitably becomes cluttered after the data set has surpassed a certain size, with overlapping photographs making viewing and finding specific photographs ever more difficult as the collection grows. Figure 1(a) exemplifies the problem by showing an unfiltered view of San Francisco photos.

Our goal is thus to facilitate a system which can automatically select representative and relevant photographs from a particular spatial region. A result of our algorithm is illustrated in Figure 1(b), where a limited set of eleven photos that were selected by our system are marked on the San Francisco map. Such collection summaries will enable users to find items more easily and browse more efficiently through large scale geo-referenced photo collections, in a manner that improves rather than degrades with the addition of more photos.

Selecting the most representative photos from a given region is a difficult task for several reasons. For instance:

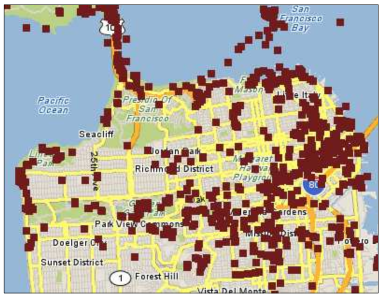
- Image analysis alone is poor at understanding the semantic content of an image, making visual relevance insufficient for summarization.
- In multi-user sets, the biases of one user's data may skew the selection towards generally insignificant subjects.
- It is difficult for an automated system to learn and assess the relevance of photos without appropriate models of human interest, as the notion of relevance is not well defined, and often subjective.


Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.


*MIR'06,* October 26–27, 2006, Santa Barbara, California, USA.


Copyright 2006 ACM 1-59593-495-2/06/0010 ...$5.00.


[1]See, for example, the ZoneTag application at http://zonetag.research.yahoo.com.




(a) All San Francisco photos


(b) An automatic summary of San Francisco photos

**Figure 1: All San Francisco photos from our dataset of 2200 geo-referenced photos, versus an automatic summary of photos, as generated by our system. One summary photo is enlarged for illustration.**

We have designed and implemented a simple algorithm that attempts to address the challenges stated above. Our algorithm utilizes metadata-based heuristics that capitalize on patterns in users' photographic behavior. Foremost among these heuristics is the premise that photographs taken at a location 'vote' for the presence of something interesting at that location.

Our algorithm considers a multitude of spatial, social and temporal metadata (such as where the photo was taken, by whom, at what time), as well as textual-topical patterns in the data, such as textual tags associated with the photo. Furthermore, the algorithm can be tuned to bias the set of results using various factors such as the social network distance of the photographers to the user making the query.

The summarization algorithm can be used in a number of applications. For example, the algorithm could be used for geographic image *search*, returning a summary of photographs from a region in response to a search query (that can be specified as a text term or a map region). In addition, the summarization can be used to assist in map-based browsing of images, for example, by selecting a subset of representative photos to show according to the map's coverage and zoom level. With or without a map, summarization can help in browsing one's photos or a group of individuals' photos to get an overview of a location or discover personally interesting areas for further exploration; automatic travel guide is a scenario that comes to mind.

Key insights from our algorithm helped us generate a new way of visualizing large collections of geo-referenced photographs. We use the techniques we developed to generate map-based tag clouds, which are described in Section 6. "Tag Maps", as we call them, can be used to visualize the contents of the collection, giving a quick overview of the textual-topical concepts that appear in the data as well as their location, importance and recency. The photos themselves are not necessarily part of the visualization. Tag Maps concepts can be applied to many other multimedia (or other) applications that exhibit patterns in text and locations.

To summarize, the contributions of this paper are:

- A new approach for generating summaries of photo collections based on geographic as well as other contextual data associated with the photographic media (Section 3).
- An outline of the requirements and the useful features for these context-based summaries (Section 3).
- An implementation of an algorithm that generates such summaries using a public set of "geo-tagged" photographs (Section 4).
- A new map-based visualization technique for photo collections that helps indicate both the important regions on the map and the textual concepts represented in those regions (Section 6).
- A proposed evaluation for geo-referenced collection summaries; we use this evaluation to compare our algorithm to several baseline methods (Section 7).

In addition, Section 5 briefly touches on potential applications. We begin by discussing the related work.

## 2. RELATED WORK

Since 2003, a number of different research efforts have considered geographic location information associated with photographs. In [19], the authors describe WWMX, a map-based system for browsing a global collection of geo-referenced photos. Several similar map-based photo browsing systems appeared on the Web in the last few years[2], most of them using "geo-tagged" images from Flickr [5] for content. All of those systems face the problem of clutter in the map interface: as the number of photos available in each location grow, the full set of images cannot possibly be shown on the map at once. While some systems default to showing the most recent photos, the WWMX system tries to handle clutter by consolidating multiple photograph markers into a single marker according to the zoom level. In our system, we avoid clutter by utilizing the additional metadata to select the best set of photographs from a region, providing potentially a better selection than the "most recent" strategy, and a more meaningful one than the "consolidation" approach.

Several projects [12, 15] use geographic data to organize photo collections in novel ways, for example, by detecting significant events and locations in a photo collection. Such structures could indeed be the basis for collection summarization. However, these projects considered personal photo collections only, and did not consider public shared pools of photos.

Looking at shared collections, some research [3, 4, 11, 14, 16] tries to use context (mostly location) information and sometimes visual features to identify landmarks in photographs. Visual analysis could be integrated in our system—once our algorithm recognizes significant locations, it can attempt to select a photo of a prominent landmark there.

[2]like http://geobloggers.com and http://mappr.com

Work in both [3, 11] considers, in a similar fashion to this work, patterns and distributions of textual terms that are associated with geo-referenced digital photos, and uses them to generate tag suggestions for new photographs. However, those projects are not designed to support collection summarization.

In the absence of location metadata, temporal metadata was also considered in the past for the purpose of photo collection summarization. In [8], Graham et al. describe an algorithm to heuristically select representative photos for a given time period in a personal collection, utilizing patterns in human photo-taking habits (later studied in [6]). Additional time-based work aims to detect events in personal collections (e.g., [2]), which could be the basis for collection summarization. However, again, all these projects considered single-photographer collections only. In public collections of timestamped photos, only when additional metadata is available (for example, the fact that all shared photos were taken in the same event), there exists the potential for time-based summaries [13].

Another possible approach for summarizing photo collections is using textual tags that are associated with the image. In Flickr [5], popular tags have pre-computed clusters of related tags. For example, the "San Francisco" tag on Flickr has several associated tag clusters[3]: "california, bridge, goldengate"; "baseball, giants, sbcpark", "deyoung, museum", "sfo, airport" and "halloween, castro". These clusters can potentially be used to generate a summary of San Francisco photos. This approach is not location-based, and the clusters often do not represent concepts that are distinct (e.g., one of Boston's clusters is "massachusetts, city, cambridge, building, architecture"). The tag clusters could possibly be used in conjunction with our method. In fact, we are using some tag-based computation to select summary photos. More directly related is a tag subsumption model [17] that can use the tag corpus to derive tags that are subsumed, for example, by the tag "San Francisco". Again, this approach can be integrated with our location-based summaries.

These projects, and others, consider various ways to alleviate the difficulties of browsing large collections of photographs, but do not provide effective ways to summarize multi-user photo collections or visualize them using maps. We believe that the potential of a geographic-based summarization method is significant, especially in conjunction with the current state of the art.

## 3. THE SUMMARIZATION APPROACH

In this section, we define the problem of summarizing a photo collection, then describe the guidelines and insights that have informed the implementation of our summarization algorithm. In Section 4 we provide the details of the algorithm.

We formalize the summarization problem as that of producing a ranking on the collection in question. In other words, we summarize a set of photos by ordering the set and selecting the top ranked photos. More formally, we are looking at the following problem: Given an album of $n$ photos, $\mathcal{A} = \{P_1, \ldots, P_n\}$, we wish to find an ordering $\omega$ of $\mathcal{A}$ such that any $k$-length prefix of $\omega(\mathcal{A})$ is the best possible $k$-element summary of $\mathcal{A}$. A summary is loosely defined as a subset that captures representativeness, relevance, and breadth in the original collection. These notions are captured through some of the following metadata attributes that are associated with the photos:

- **Location.** Photo $P_i$ was taken at location $(x_i, y_i)$.[4]
- **Time.** Photo $P_i$ was taken at time $t_i$.
- **Photographer.** Photo $P_i$ was taken by user $u_i$.
- **Tags.** Photo $P_i$ was manually assigned the list of tags (i.e., textual labels) $w_i$.
- **Quality.** Photo $P_i$ is associated with an externally derived parameter $q_i$ that represents its quality.
- **Relevance.** Photo $P_i$ is associated with a relevance factor $r_i$. Relevance can include arbitrary bias based on parameters such as recency, the time of day, the day of the week, the social network of the user, user attributes, and so forth.

Note that The relevance attribute can introduce subjectivity, allowing us, for example, to tune the results to the user who is making the query and the context of the query.

While there is no accurate formal model for what constitutes a "good" summary of a collection of geo-referenced photographs, we follow a few simple heuristics that try to model and capture human attention, as reflected in the set of photos taken in a region. Among these heuristics are the notions that:

- Photographs are taken at locations that provide views of some important object or landmark.
- A location is more relevant if the photos around it were taken by a large number of distinct photographers.
- If available, location-based patterns of textual tags can reflect the presence of an important landmarks in a location.

In addition to the heuristics listed above, a desired summary would also (a) represent a broad range of subjects, instead of thoroughly displaying a few, and (b) allow personal or query bias to modify the algorithm's results.

In the next section we describe the summarization algorithm that we developed based on these guidelines.

## 4. ALGORITHM FOR SUMMARIZATION

As described in Section 3, our summarization algorithm produces a ranking of the photos in the collection; each prefix of this ranking can serve as a collection summary of the corresponding size. Producing this ranking is a two-step process, a clustering step followed by a ranking step on the resulting clustering hierarchy. In particular:

1. We apply a modified version of the Hungarian clustering algorithm [7] to our collection of photographs. This algorithm receives the photograph locations as an input, and organizes them into a hierarchical clustered structure.
2. We compute a score for each cluster in the hierarchy.
3. Finally, we generate a flat ordering of all photos in the dataset by recursively ranking the sub-clusters at each level, starting from the leaf clusters, and ending at the root.

[3]http://flickr.com/photos/tags/sanfrancisco/clusters/

[4]Notice that this 'photo origin location' is different than the 'target location', the location of the photographed object.

Note that while the clustering is a fixed one-time computation, the ranking step can be re-evaluated, allowing users to specify a personal bias or preference towards any of the metadata features. Alternatively, the ranking can also be modified to utilize implicit bias in the query context (e.g., the identity of the user making the query).

To illustrate the process and the scoring mechanism we use a hypothetical example, presented in Figure 2. In this figure, a leaf node represents a single photograph, annotated with the identity of the photographer and a single textual tag (in practice, of course, more tags can be associated with each photo). The tree represents the hierarchy created by the clustering algorithm.

Next, we describe the algorithm in detail. First, we discuss the clustering algorithm that produces the clustering hierarchy. Then, we describe how to produce a ranking of all photos in a single node of the above mentioned clustering hierarchy, assuming that all nodes in the hierarchy are associated with scores. Finally, we show how we can generate such scores for the nodes in the hierarchy.

## 4.1 Clustering

Our method requires a hierarchical clustering algorithm; as noted above, we use the Hungarian clustering algorithm [7]. This algorithm identifies a hierarchy of clusters within a given dataset of $n$ points, based only on the distances between those points. In our system, the input to the clustering algorithm is a set of points in the plane, representing the locations of the photographs,[5]

$$\mathcal{A} = \{(x_i, y_i) \in \mathbb{R}^2,\ 1 \le i \le n\}\ . \quad (1)$$

The output is a clustering of these photo locations, $C(\mathcal{A})$, where $C(\mathcal{A})$ is a tree. Each node in the tree represents a subset of $\mathcal{A}$, the root of the tree represents the entire set, the children of each node are a partition (or clustering) of the subset that is associated with that parent node, and the leaves of the tree are the points in $\mathcal{A}$.

The classical Hungarian method is an efficient algorithm for solving the problem of minimal-weight cycle cover. In that problem, one is given a weighted graph and needs to find a cover of that graph by disjoint cycles with minimal total weight. This algorithm serves as the basic building block for a clustering method that is dubbed *The Hungarian clustering method*. Viewing $\mathcal{A}$ as a complete weighted graph, where the weight of each edge is the Euclidean (geographic, in this case) distance between the two nodes that it connects, the Hungarian clustering method subjects that graph to the classical Hungarian method. The disjoint cycles, produced by the Hungarian method, are viewed as a partition of the data-set. The clustering algorithm then proceeds by hierarchical merging of the disjoint cycles, until the produced clusters are perceived as complete clusters (through some "completeness" criteria) and then the hierarchical merging stops. We use the Hungarian Clustering algorithm because of two features that it boasts: It is an hierarchical clustering algorithm, and it does not depend on the number of clusters as an input.

The clustering hierarchy $C(\mathcal{A})$ is used to create a ranking of all photos. In order to describe the ranking algorithm, let us first assume that the nodes in the hierarchy have been assigned a score that embodies the importance of the cluster of photos that corresponds to that node.

[5]For convenience, we use the same notation, $\mathcal{A}$, to denote the photo set as well as the set of photo locations.

## 4.2 Ranking Framework

Given a hierarchical clustering $C(\mathcal{A})$ on the locations of all photographs, and a score for every node (cluster) in that hierarchy, our goal is then to produce a ranking of all items in the collection. We describe a recursive interleaving algorithm that uses the clustered structure and the corresponding scores in order to produce a natural flat ordering. In the next section we outline a way to generate the scores.

Going bottom up, the ranking algorithm considers each node $\mathcal{B}$ in the hierarchy $C(\mathcal{A})$ and outputs an ordering $\omega(\mathcal{B})$ that represents a ranking of photos in $\mathcal{B}$. Finally, when executing on the root node that corresponds to the entire set $\mathcal{A}$, we get the ordered sequence, $S := \omega(\mathcal{A})$, that describes a ranking of all photos in $\mathcal{A}$. Applying this algorithm to the example in Figure 2, a possible output could be the ranking $S = (6, 8, 4, 5, 7)$, where the numbers in the sequence correspond to the numerals of the leaves in the tree in Figure 2.

For simplicity of notations, we describe the action of the algorithm on the root node, $\mathcal{A}$. Actions on other nodes are performed in the same manner. We assume that we identified $m$ sub-clusters in $\mathcal{A}$, $\mathcal{A} = \bigcup_{i=1}^{m} \mathcal{A}_i$; namely, node $\mathcal{A}$ has $m$ direct descendents. In addition, assume that the photos in each sub-cluster of $\mathcal{A}$ have been ranked recursively according to this algorithm, and that each of the nodes $A_i$ is associated with some score $s(\mathcal{A}_i)$ such that (without loss of generality)

$$s(\mathcal{A}_1) \ge s(\mathcal{A}_2) \ge \cdots \ge s(\mathcal{A}_m)\ . \quad (2)$$

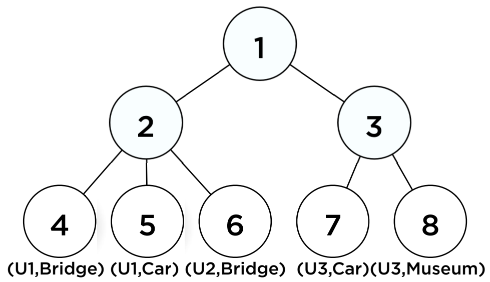



**Figure 2: Hypothetical clustering hierarchy; the leaves are photos, each associated with a user and a single tag.**

Our goal is to produce a ranking that would balance the contradicting properties of depth and breadth of coverage. In the field of Information Retrieval, some measures are used to balance results in terms of relevance (depth) and broadness (breadth) [1, 10, 20]; for various reasons, these measures are not applicable here. For our problem, depth requires that the photos in a cluster are selected from sub-clusters roughly according to the ratio of their scores. For example, consider the second level of the hierarchy in Figure 2, which consists of two clusters, denoted by $\mathcal{C}_2$ and $\mathcal{C}_3$, and assume that $s(\mathcal{C}_2) : s(\mathcal{C}_3) = 5 : 3$. We would like to interleave the photos from these two clusters so that in any section of the sequence $S$, the frequencies of photos from the two clusters relate to each other as closely as possible to their score ratio in the whole dataset, i.e., $5 : 3$. On the other hand, breadth requires that each sub-cluster should be represented to some extent early in the ranking of its parent cluster.

In order to attain some amount of depth, breadth, and consistency, we interleave photos from sub-clusters in the following manner. The ordered sequence of photos for $\mathcal{A}$ will have two parts: a short *header* $\mathcal{H}$ followed by a *trailer* $\mathcal{T}$, where $S(\mathcal{A}) = \mathcal{H}||\mathcal{T}$.

The header $\mathcal{H}$ will include a photo from all prominent sub-clusters. To that end, we define a threshold $0 < w < 1$, and then a cluster $\mathcal{A}_i$ is deemed prominent if

$$\frac{s(\mathcal{A}_i)}{\sum_{j=1}^{m} s(\mathcal{A}_j)} \geq w \ .$$

Assume that there are $m'$ prominent sub-clusters among the $m$ sub-clusters, with $0 \leq m' \leq m$. Then in view of assumption (2), the header is

$$\mathcal{H} = (\mathcal{A}_{1,1}, \mathcal{A}_{2,1} \cdots, \mathcal{A}_{m',1}) \ ,$$

where $\mathcal{A}_{i,1}$ is the most relevant photo from cluster $\mathcal{A}_i$. This header is then followed by a trailer, $\mathcal{T}$. In order to generate the trailer, we first remove from each sub-cluster the photo that was selected for the header, recalculate the sub-cluster scores, and then assign each sub-cluster a probability that equals its score divided by the sum of scores of all sub-clusters. Those probabilities are then used to randomly select a sub-cluster. If sub-cluster $\mathcal{A}_i$ was selected, we remove its top-ranked photo, append it to $\mathcal{T}$ and repeat, until all photos have been selected.

By now we have described how to generate the cluster hierarchy and produce a ranking on the photos in that hierarchy, under the assumption that all nodes are associated with scores. We therefore proceed to describe a key aspect of the algorithm: the computation of the scores for a given cluster (node).

## 4.3 Scoring Clusters

The score of a cluster $\mathcal{A}_i$ depends on several factors, including the following:

1. The sum of relevance factors (see Section 3) of all photos in the cluster,
$$\rho_i = \sum_{P_j \in \mathcal{A}_i} r_j \ .$$
2. The tag-distinguishability of the cluster, $\tau_i$ (explained below).
3. The photographer-distinguishability of the cluster, $\phi_i$ (explained below).
4. The density of the cluster. More specifically, let $\sigma_{x,i}$ and $\sigma_{y,i}$ denote the standard deviation of the $x$ and $y$ coordinates, respectively, of all points in $\mathcal{A}_i$, and let
$$\sigma_i = \left((\sigma_{x,i})^2 + (\sigma_{y,i})^2\right)^{1/2} \ .$$
We define the cluster density as
$$\delta_i = 1/(1 + \sigma_i) \ . \tag{3}$$
5. The sum of image qualities (see Section 3) of all photos in the cluster,
$$\kappa_i = \sum_{P_j \in \mathcal{A}_i} q_j \ .$$

While most of the above factors are derived only from data that is contained in the photo collection, the relevance factor can introduce bias by subjective requirements. The relevance factor $r_i$ of a photo $P_i$ can incorporate parameters such as recency, the time of day, the time of the week, the identity of the photographer, etc. These parameters can be specified by a user making the query, or set by the system according to the application or the query context. Each photo is assigned a score $\theta(P_i)$ in the range $[0, 1]$ for each such parameter. The final relevance score, $r_i$, may be a weighted average of all those parameter scores.

The two interesting factors in the score computation are the tag- and photographer-distinguishability scores of clusters. These values are intended to represent how strongly a particular cluster is associated with specific tags or photographers.

### 4.3.1 *Tag-distinguishability of clusters*

Tag-distinguishability aims at detecting distinct or unique concepts that are represented in a given cluster, as those may indicate the presence of some interesting landmarks or objects in that cluster. For example, in Figure 2, the tag "bridge" appears in two photos from Cluster $\mathcal{C}_2$, and does not appear elsewhere. As a consequence, $\mathcal{C}_2$'s score improves. On the other hand, the tag "car" appears in photos from both $\mathcal{C}_2$ and $\mathcal{C}_3$ and therefore does not help to distinguish either of them.

Formally, each photo $P_j$, $1 \leq j \leq n$, is tagged with tags that are drawn from a finite dictionary, $T$. Hence, tagging may be viewed as a mapping $P_j \mapsto T(P_j) \subset T$. For all $t \in T$ and $1 \leq i \leq m$, let

$$\mathtt{tf}_{t,i} = \frac{|\{P_j \in \mathcal{A}_i : t \in T(P_j)\}|}{|\mathcal{A}_i|} \tag{4}$$

denote the relative frequency of the tag $t$ in $\mathcal{A}_i$, (or *term frequency* as it is referred to in Information Retrieval). We often found that this measure biases towards tags that have been used frequently by one user in the same cluster. An alternative frequency calculation can be based on the fraction of photographers in this cluster that have used the tag $t$:

$$\mathtt{tf}_{t,i} = \frac{|\{u \in U_i : t \in T(P_j), P_j \in \mathcal{A}_i, P_j \in B_u\}|}{|U_i|} \tag{5}$$

where $U_i$ is the set of users that have taken photos in cluster $\mathcal{A}_i$, and $B_u$ is a set of photos taken by user $u$.

We also use the *inverse document frequency*, which is a measure of the overall frequency of the tag $t$ in the entire photo collection,

$$\mathtt{idf}_t = \frac{n}{|\{P_j \in \mathcal{A} : t \in T(P_j)\}|} \ . \tag{6}$$

There are several ways to combine these two scores to measure how the term $t$ distinguishes the cluster $\mathcal{A}_i$ from other clusters. Let us denote such measures by $\tau_{t,i}$. The usual measure in Information Retrieval is the `tf-idf` weight (term frequency – inverse document frequency). That measure is defined as

$$\tau_{t,i} := \mathtt{tfidf}_{t,i} = \mathtt{tf}_{t,i} \cdot \mathtt{idf}_t \ . \tag{7}$$

Another alternative to (7) which is used in Information Retrieval is

$$\tau_{t,i} := \mathtt{tfidf}_{t,i} = \mathtt{tf}_{t,i} \cdot \log(\mathtt{idf}_t) \ . \tag{8}$$

In both cases, large values of $\tau_{t,i}$ indicate that the number of occurrences of $t$ in $\mathcal{A}_i$ is large with respect to its number of occurrences elsewhere.

We would like to note that in the usual `idf` weight, the inverse document-frequency involves the number of clusters in which the tag appears, as opposed to the total number of actual tag occurrences, as given in (6). However, the usual definition is not suitable for cases where the number of clusters (documents) is small. In such cases, a single random occurrence of a tag in a cluster has a significant effect on the usual measure, while in the alternate approach we opted for it would be hardly noticeable.

Next, we need to define an overall tag-distinguishability measure for $\mathcal{A}_i$, denoted $\tau_i$, based on the tag-distinguishability measures of all tags in the cluster. We compute the overall score by using the Euclidean measure based on the $\ell_2$-norm,

$$\tau_i = \left( \sum_{t \in T} \tau_{t,i}^2 \right)^{1/2} . \tag{9}$$

We directly evaluate the effectiveness of our approach to "tag scoring" in Section 7.

#### 4.3.2 Photographer-distinguishability of clusters

The measure of photographer-distinguishability (or user-distinguishability) is, roughly, inversely correlated to the number of photographers associated with a given cluster. The fewer active photographers in a cluster, the lower the likelihood the cluster will be semantically meaningful. For example, in Figure 2, all the photos in Cluster $\mathcal{C}_3$ were taken by the same user ($U3$), while that user did not take any photos elsewhere. Consequently, the cluster seems to have less general appeal than $\mathcal{C}_2$.

Hence, much like for tags, we consider a `tf-idf`-like score for the correlation between a cluster $\mathcal{A}_i$ and a photographer $u$. Let $\mathcal{B}_u$ denote the set of photos that were taken by the photographer $u$. Then the score is given by

$$\phi_{u,i} := \mathtt{tf}_{u,i} \cdot \mathtt{idf}_u \tag{10}$$

where

$$\mathtt{tf}_{u,i} = \frac{|\mathcal{A}_i \cap \mathcal{B}_u|}{|\mathcal{A}_i|} \tag{11}$$

is the relative portion of photographer $u$ in photos from cluster $\mathcal{A}_i$, and

$$\mathtt{idf}_u = \frac{n}{|\mathcal{B}_u|} \tag{12}$$

is the inverse of the photographer's relative portion in photos from the entire dataset. Note that (10), (11) and (12) are equivalent to (7), (5) and (6), respectively. As previously, compare (8) with (7), we may replace (10) with

$$\phi_{u,i} := \mathtt{tf}_{u,i} \cdot \log(\mathtt{idf}_u) . \tag{13}$$

Finally, the overall photographer-distinguishability is defined as

$$\phi_i = \left( \sum_{u} \phi_{u,i}^2 \right)^{1/2} . \tag{14}$$

According to the guidelines in Section 3, while large tag-distinguishabilities should contribute towards an increase in a cluster's score, the photographer-distinguishability should have an opposite effect. The more a given cluster is associated with a single photographer (or few photographers), the less we are interested in that cluster.

Next, we describe how to merge all these factors into a single score for each cluster.

#### 4.3.3 Overall Cluster Score

The score $s(\mathcal{A}_i)$ of the cluster $\mathcal{A}_i$ should depend in a monotonically increasing manner on the relevance factor, $\rho_i$, and the image quality factor, $\kappa_i$. The score should also depend in a monotonically increasing manner on the density measure of the cluster, $\delta_i$ (3), and on $\tau_i$, the tag-distinguishability measure of the cluster. Finally, the score must depend in a monotonically decreasing manner on $\phi_i$, the photographer-distinguishability measure of the cluster, as discussed above. Therefore, the overall score is:

$$s(\mathcal{A}_i) = h(\kappa_i, \delta_i, \tau_i, \phi_i^{-1}) \cdot \rho_i \tag{15}$$

where $h$ could be, typically, a geometric mean or a weighted average of its variables, and the weights may be chosen and fine-tuned by experimentation.

### 4.4 Final Considerations

We have described in this section the full framework of our algorithm: how the clustering is done, how scores are computed for each node in the cluster hierarchy, and how an ordering is produced on all photos given the scores and the cluster hierarchy. In practice, we found that it was necessary to prevent clusters from subdividing past some minimum size. Computing a ranking for a small cluster is meaningless. For example, there is not enough information (photos in the clusters) to compute a relevant tag- or photographer-distinguishability score. In order to solve this problem, we simply enforced a minimum size on all non-leaf clusters, by merging nodes at the lowest levels of the hierarchy.

The way ranking is performed for these flat "edge" clusters is simple, yet different than the generalized method described above. The system computes the top-scoring tags for each flat cluster using the tag-distinguishability method, and then picks a photo with tags that best match these top tags. For example, in Figure 2, photo 6 is likely to be picked first because of the tag 'bridge', which is associated with that photo, and also appears often in $\mathcal{C}_2$ and rarely elsewhere. Other approaches for ranking photos for the flat leaf clusters may be choosing photos that maximize the visual similarity to other photos in this cluster, or some combination of such tag- and image-based similarity, as well as quality and relevance factors associated with the photos. Joshi et at [9] propose, in a different context, a solution that can be applied here.

## 5. SAMPLE APPLICATION

The summarization algorithm has a number of possible uses. As mentioned above, the summaries could be used to support "semantic zoom" on large collections of digital photographs, or help in browsing/searching a large collection by showing just the summarized results. This latter application scenario is partially tested in our evaluation, Section 7.

### 5.1 Map-Based Browsing with Semantic Zoom

When location data is available, a user may wish to view photographs placed on a map at the locations at which they were taken. This can be an excellent way to view and understand photos in context. Unfortunately, after the data set has surpassed a certain size, the space inevitably becomes cluttered, and the overlap of photos makes comprehending the full dataset impossible. We believe that by using semantic zoom techniques that are based on our summarization algorithm, map-based photo browsing can be a practical reality.

Semantic zoom is the concept that zooming through some space in which digital objects are embedded, such as a map or a timeline, should be accompanied by a corresponding shift in the quantity of content presented. In our case, this means presenting to the user a subset of the photographs, where the size of the subset is appropriate to the current zoom level. The images in the subset are chosen according to the rank given to them by our algorithm. As the user zooms in, more photographs (that were ranked lower) are revealed, thus bringing the content into more detail. At any zoom level, panned to any region, the user should see a small set of photos that best represents that region. Given an ordering on the photos, the implementation of this interaction becomes trivial. When viewing any region, display the $k$ best photos that were taken within that region. When the user zooms in further, our algorithm only needs to go down the ranked list of photos, and add photos to the map until the currently viewed map region is populated with the right amount of photos (as determined by the application).

## 6. VISUALIZING COLLECTIONS: TAG MAPS

*Tag Maps* is a visualization we developed to expose textual topics that are tied to a specific location on a map. The tags that are deemed relevant can be shown at the location where they "occur", and displayed in a size that corresponds to the tag's importance, as shown in Figure 3. Tag Maps have some interesting parallels with *tag clouds* (that are, essentially, lists of tags with font sizes that are weighted by the tags' popularity), which have recently been commonplace in various social-media websites.

While Tag Maps are a generic form of visualization, our summarization algorithm can be used to seed a Tag Map visualization for geo-referenced photo collections. Rather than display representative photographs at their respective locations, it is possible to convey the concepts represented in the dataset through the tags. The visualization is based on the textual tags that are associated with the photos data, and mainly uses the tag-distinguishability scores, along with the clustering. The visualization can also use the other factors that inform our algorithm's results.

In our algorithm, we calculate a tag-distinguishability score $\tau_i$ for every cluster $\mathcal{A}_i$ at every level of the hierarchy. In the process, an individual score $\tau_{t,i}$ is calculated for each tag $t$ as described in Section 4. This tag score, as a variant of `tf-idf`, can be thought of as a measure for how well and how *uniquely* $t$ represents $\mathcal{A}_i$.

Consider the following mapping: for some 'natural' level of the hierarchy, (to be determined by the clustering algorithm), and for each cluster $\mathcal{A}_i$ in that level, pick the tag $t \in T$ that maximizes $\tau_{t,i}$. We thus have a set of clusters, and one tag to represent each cluster, with a score associated with that cluster and tag.

The chosen tags are displayed on the map directly above the centroid of their respective cluster. The displayed size of the tag corresponds to its computed score. Notice that the displayed size can also reflect other factors that go into the cluster score such as relevance, density, photographer-distinguishability and so forth. Another dimension of information can be easily encoded in the Tag Map; for example, the gray level of the tag can represent recency – recent tags are darker, older tags are lighter.

Figure 3 shows the Tag Maps produced by our algorithm to represent photo collection in San Francisco and London. In both cases, the photo collections used to generate this visualization are 'geo-tagged' images from Flickr (see Section 7). Over 1000 such geo-tagged photos were used for each city. Indeed, relevant concepts represented in these cities arise from the visualization: Lombard Street, Golden Gate Bridge, Golden Gate Park, Twin Peak in the San Francisco map; Buckingham Palace, Big Ben, Hyde Park, and more in the London map. Notice that, as discussed in Section 3, the tags in our Tag Maps represent "photo spots" and not necessarily the locations of the objects themselves (see for example the "Golden Gate Bridge" tags that appear in two different view points of Golden Gate Bridge). Also note that our results were not free of problems. For example, Figure 3(a) shows the appropriate but not representative tag "trees" for a location inside one of the parks (left side of the map). On the other hand, Figure 3(b) shows a number of clusters tagged "London", which our algorithm should have scored lower given that "London" must appear in most clusters in large numbers.


(a) San Francisco


(b) London

**Figure 3: Tag Maps of San Francisco and London**

Finally, Tag Maps can be different at various zoom levels, displaying more tags from more clusters as the map interface is zoomed in. Moreover, Tag Maps can be used for collections other than geo-tagged photographs: e.g., visualizing popular search keywords from different areas on the map. We evaluate Tag Maps, and through it key aspects of our algorithm, in the next section.

# 7. EVALUATION

We implemented a version of our algorithm, and performed a number of user evaluations to compare our system with a number of baseline approaches. Our current summarization implementation utilizes only three features: location, photographer, and tags. The system clusters the input photo locations using the Hungarian clustering algorithm, and then computes a score for each cluster $\mathcal{A}_i$ at each level, by taking the product of (a) the number of photos in $\mathcal{A}_i$, (b) the tag-distinguishability of $\mathcal{A}_i$, and (c) the photographer-distinguishability of $\mathcal{A}_i$. Given these scores, we set the header/trailer threshold such that 5% of the photos are added to the header, and sample the rest randomly as the trailer, as described in Section 4.2.

For all our tests, we used photos from the same pool of images, 'geo-tagged' photos from the popular photo sharing website Flickr [5]. Geo-tagged images are photos that are associated with latitude and longitude tags, often (but not always) representing the exact location where the photo was taken. We retrieved all such photos from the San Francisco area; there were over 2200 such geo-tagged San Francisco photos on Flickr. We refer below to this dataset as $P_{eval}$.

## 7.1 Evaluation Framework

Today, the size and (more so) geographic distribution of the currently available geo-tagged datasets do not allow for a task-based evaluation of the system, such as measuring its usefulness in browsing. Instead, we performed direct evaluation of our system, having users judge its output. The goals of our evaluation were to:

- Verify that our algorithm scores for tag- and photographer-distinguishability are meaningful and accurate.
- Determine whether the summary algorithm identifies representative locations in a given spatial region better than baseline methods.
- Test whether the photos selected as a summary by our algorithm form a better summary than photos chosen by various baseline methods.

These goals are clearly dependent on subjective measures; we therefore performed our evaluation by user tests. We executed three different experiments to accomplish these goals. These tests are listed next.

## 7.2 Tag Maps Test

The goal of the Tag Maps test was to determine if the tag-distinguishability features of the algorithm are useful and meaningful, in that (a) important textual concepts that are related to specific locations are surfaced and (b) unimportant or highly personal tags are demoted. We used our tag maps visualization (see Section 6) for this test. The selection process of the visualization is somewhat different than that of our algorithm, selecting one prominent tag for each cluster instead of scoring the cluster according to all tags in it; however, the selection, location and displayed size of each tag is directly related to our algorithm‘s cluster-scoring mechanism.

For the purpose of evaluation, we performed a within-subject experiment. We showed the subjects tag maps that were based on our clustering results. For each cluster, the top-scoring tag was selected and shown according to one of three tag scoring variations: either (a) the basic `tf-idf` score of that tag in the cluster, as described earlier in (4) , (b) the same with a threshold in which clusters containing photographs by only one user are not displayed, and (c) the alternative tag scoring method determined by the fraction of *users* who used the tag in this cluster, described by (5), with the same threshold as (b). This latter option was used to generate the tag map shown in Figure 3. In all cases, the displayed size of the tag was proportional to its score.

We asked each subject to rate the three different tag maps in terms of (1) whether the tags appear in an appropriate location, (2) how representative of each location is its displayed tag, (3) how well the size of tags represents their importance in the displayed map region and (4) overall, how well the map represents the region.

Due to lack of space, we do not provide complete results for this test. To summarize, the results indicated that the first two methods surfaced many irrelevant tags; the size of the tags was often not reflective of the importance of the tag to the region. Omitting tags for clusters with a single user reduced clutter and removed many inappropriately large tags. Switching our scoring method to (c), that is based on counting the fraction of photographers using a given tag, brought significant improvements. Subjects also rated this scoring scheme higher in terms of the quality of the overall representation of the map region. This initial finding both determined our implementation technique for the Tag Map visualization, and suggested that photographer-distinguishability is an important factor in generating collection summaries.

## 7.3 Location Importance Test

We executed the Location Importance test in order to verify that our algorithm selects a good subset of locations for a summary of a region. In this test, we examine the location of photos selected, rather than the content of the photos. For example, Figure 1(b) shows that our algorithm summarized $P_{eval}$ by selecting photos from a viewpoint of the Golden Gate Bridge, in Golden Gate Park, at the famous Lombard Street, and more – arguably a good set of locations for a summary of San Francisco locations.

For this test (and the next one), we compared four basic conditions that represent different ways to generate summary sets from a collection of geo-tagged images. These conditions include our summarization algorithm, "Interestingness", "most recent", and "random". The conditions, and a short description of each, are listed in Table 1. As the last two conditions are least stable, and are likely to change every time such a selection is made, we have tried two instances of each condition (i.e., most recent photos at two different points in time, and two different random sets), for a total of six experimental conditions. Notice that all conditions are naturally biased towards selecting photos from more popular regions, simply due to photo density: more photos are taken in popular locations, and therefore the probability that photos from these locations will be selected is higher.

To find out if the locations of the photos selected in the different experimental conditions were meaningful, we first had to compile a "ground truth" list of the interesting locations in San Francisco. To this end, we asked 25 people to each write down a list of $5 - 10$ top tourist locations in San Francisco. We compiled their answers and ranked the locations by the number of times each was mentioned. We selected the top 15 locations, which were chosen by at least

**Table 1: Basic Experimental Conditions for the Location and Summary Tests**

| *Condition* | *Description* |
|---|---|
| Summarization | Our Algorithm results on $P_{eval}$ |
| Interestingness | Photos from $P_{eval}$ with the highest Flickr "interestingness" score[a] |
| Recent | The most recent photos from $P_{eval}$, with no more than one photo per user. |
| Random | A random selection of photos from $P_{eval}$ |

[a]Flickr interestingness is the website's measure of the attention given by other users to a photo.

3 people, for our test. As a second ground truth list, we have used the top 10 locations from Yahoo! Travel's "Things to do in San Francisco". As expected, both lists included Fisherman's Warf (Pier 39), Golden Gate Bridge and Alcatraz.

We checked how many of the top 10 photos' locations, as selected by each experimental condition, matched (in terms of location) the ground truth lists. A positive match was awarded when the location matched, even if the photo did not portray the actual attraction (the test we performed in order to verify the content of the photos is covered in the next section). For example, if a photo that was taken in Alcatraz was selected for the summary, it matched both our ground truth lists.

#### 7.3.1 Results

Figure 4 shows the percentage of summary photos from each condition that were taken in a location listed in one of our ground truth location lists. For example, 60% of the locations in the summary generated by our algorithm appeared in our subjects' top 15; 50% of the locations appeared on Yahoo's guide top 10 list. Our algorithm clearly performed better than all other conditions, for both ground truth lists.

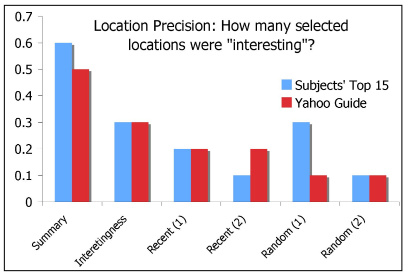


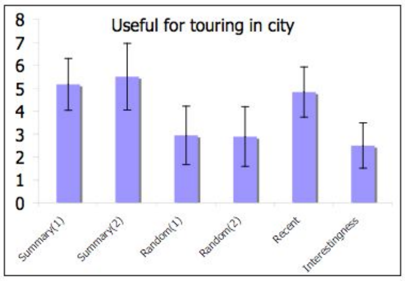

**Figure 4: Loction Precision: the number of photos in each condition's summary that matched a location that appears in ground truth lists.**

### 7.4 Summary Relevance Test

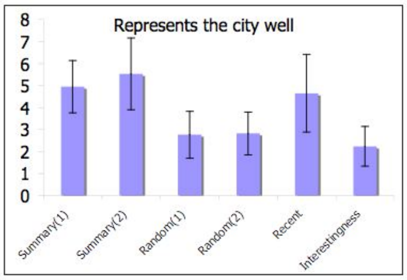
The purpose of the Summary Relevance test was to directly evaluate the set of photographs that were selected by our algorithm as the summary of dataset $P_{eval}$, comparing against the other base methods. In this test, we showed our subjects summaries consisting of nine photos from the dataset. Each of these summaries was generated by one of the experimental conditions listed in Table 1: our summarization algorithm, and the three other base methods. We performed an within-subject evaluation with a set of 18 subjects. Each subject was shown the nine photos selected by each summary, and was asked to rate each such summary on various criteria: relevance to the city, attractiveness of the photos, usefulness for showing the city to a friend, and the extent to which the entire city is represented.

#### 7.4.1 Results

Figure 5 shows a summary of the results. We show subjects' ratings for each condition in response to two selected questions from the survey (trends were similar for all questions). The selections made by the two variations of our algorithm, Summary (1) and Summary (2), were better than the Random and Interestingness selections. The Recent selection used for this test happened to be a good selection of the city. However, it is hard to imagine that such a selection could be consistently representative across time. In fact, the previous Location Importance Test featured two other 'recent' selections that performed quite badly.




(a) "Is this set useful for showing San Francisco to a friend?"




(b) "How well does this set represent San Francisco?"

**Figure 5: Subjects' ratings for two questions from the experiment.**

To summarize, the three tests we performed had shown that the summarization algorithm performs quite well in identifying important photographic locations, and selecting the actual photos for the summary. The summary algorithm's performance exceeded the three baseline methods that have proven both less effective, and less robust in the face of changing data.

## 8. CONCLUSION AND FUTURE WORK

We believe that our approach to selecting representative photos from geo-referenced collections is useful for many applications involving large collections of geo-referenced digital photographs. We have shown a way to generate such representative summaries, and how to generate Tag Maps visualization of these datasets. A direct evaluation of our algorithm resulted in a favorable outcome.

We found that some aspects of the system need to be improved. In particular, while the system often identified the important locations where representative photos are likely to be found, extracting visual features from the clusters could potentially assist in selecting better photos for the final summary. Using such technology for place and landmark recognition [4, 16], augmented by our tag-based selection, and given that a location was already identified, may be critical in ensuring that a "most representative" image is selected.

An interesting question is the information requirements of our algorithm. How many photos are needed in a given region, and from how many photographers, before meaningful results are available? For now, we can only attest to the fact that the algorithm tested well for city-sized regions with roughly 1000 photos or more. How to adjust the system parameters to operate on different densities of photos?

In addition, at this point our dataset was not large enough to study the proposed biasing mechanism (for example, biasing for recent photos or photos from a user's social network). We would like to explore that further in our future work. Other possible paths could be trying to correlate gazetteer data, or textual data from other sources (e.g., subsumed tags [17]) with tags derived by our algorithm.

The phenomenal growth of personal and shared digital photo collections presents considerable challenges in building navigation and summarization applications. By utilizing our summarization algorithm, which can be parameterized by user and contextual bias, we enable users to view the most relevant samples from large-scale geo-referenced photo collections, with little to no effort.

## 9. REFERENCES


[1] Jaime Carbonell and Jade Goldstein. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *SIGIR '98: Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 335–336, New York, NY, USA, 1998. ACM Press.

[2] Matthew Cooper, Jonathan Foote, Andreas Girgensohn, and Lynn Wilcox. Temporal event clustering for digital photo collections. In *Proceedings of the eleventh ACM international conference on Multimedia*, pages 364–373. ACM Press, 2003.

[3] Marc Davis, Simon King, Nathan Good, and Risto Sarvas. From context to content: leveraging context to infer media metadata. In *Proceedings of the 12th International Conference on Multimedia (MM2004)*, pages 188–195. ACM Press, 2004.

[4] Marc Davis, Michael Smith, Fred Stentiford, Adetokunbo Bambidele, John Canny, Nathan Good, Simon King, and Rajkumar Janakiraman. Using context and similarity for face and location identification. In *Proceedings of the IS&T/SPIE 18th Annual Symposium on Electronic Imaging Science and Technology*, 2006.

[5] Flickr.com. http://www.flickr.com.

[6] Ullas Gargi. Consumer media capture: Time-based analysis and event clustering. Technical Report HPL-2003-165, HP Laboratories, August 2003.

[7] Jacob Goldberger and Tamir Tassa. The hungarian clustering method. Technical report, 2006. Submitted for publication.

[8] Adrian Graham, Hector Garcia-Molina, Andreas Paepcke, and Terry Winograd. Time as essence for photo browsing through personal digital libraries. In *Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries*, 2002.

[9] Dhiraj Joshi, James Z. Wang, and Jia Li. The story picturing engine: finding elite images to illustrate a story using mutual reinforcement. In *MIR '04: Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval*, pages 119–126, New York, NY, USA, 2004. ACM Press.

[10] John Lafferty and Chengxiang Zhai. Document language models, query models, and risk minimization for information retrieval. In *SIGIR '01: Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 111–119, New York, NY, USA, 2001. ACM Press.

[11] Mor Naaman, Andreas Paepcke, and Hector Garcia-Molina. From where to what: Metadata sharing for digital photographs with geographic coordinates. In *10th International Conference on Cooperative Information Systems (CoopIS)*, 2003.

[12] Mor Naaman, Yee Jiun Song, Andreas Paepcke, and Hector Garcia-Molina. Automatic organization for digital photographs with geographic coordinates. In *Proceedings of the Fourth ACM/IEEE-CS Joint Conference on Digital Libraries*, 2004.

[13] Rahul Nair, Nicholas Reid, and Marc Davis. Photo LOI: Browsing multi-user photo collections. In *Proceedings of the 13th International Conference on Multimedia (MM2005)*. ACM Press, 2005.

[14] Neil O'Hare, Cathal Gurrin, Gareth J.F. Jones, and Alan F. Smeaton. Combination of content analysis and context features for digital photograph retrieval. In *2nd IEE European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies*, 2005.

[15] A. Pigeau and M. Gelgon. Organizing a personal image collection with statistical model-based ICL clustering on spatio-temporal camera phone meta-data. *Journal of Visual Communication and Image Representation*, 15(3):425–445, September 2004.

[16] Aran Qamra, C. Tsai, and Edward Y. Chang. A scalable system for landmark recognition in digital photographs. Technical report, UCSB, 2005.

[17] Patrick Schmitz. Inducing ontology from flickr tags. In *Proceedings of the Workshop on Collaborative Web Tagging*, 2006.

[18] Susan Sontag. *On Photography*. Picador, New York, NY, 1977.

[19] Kentaro Toyama, Ron Logan, and Asta Roseway. Geographic location tags on digital images. In *Proceedings of the 11th International Conference on Multimedia (MM2003)*, pages 156–166. ACM Press, 2003.

[20] Hal R. Varian. Economics and search. *SIGIR Forum*, 33(1):1–5, 1999.