In the Working Notes of the 1995 AAAI Fall Symposium on AI Applications in Knowledge Navigation and Retrieval.

Scatter/Gather as a Tool for the Navigation of Retrieval Results

Marti A. Hearst*   David R. Karger†   Jan O. Pedersen*

*Xerox Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304, USA
{hearst,pedersen}@parc.xerox.com

†Laboratory of Computer Science
Massachusetts Institute of Technology
Cambridge, MA 02139
karger@theory.lcs.mit.edu

Abstract

An important information access problem arises when the user is confronted with a very large number of documents that have been retrieved in response to a query. In this paper we explore the use of a technique, called Scatter/Gather, for the navigation of large collections of retrieved documents. Scatter/Gather clusters the documents into semantically coherent groups on-the-fly and presents descriptive summaries of the groups to the user. These groups can be used in several ways: to identify useful subsets of documents to be perused with other tools, to eliminate subsets whose contents appear nonrelevant, or to select promising document subsets for reclustering into more refined groups. This paper describes the Scatter/Gather algorithm and illustrates its application to retrieval results via two examples.

Introduction

If a user of an information access system issues a query that retrieves a very large number of documents, that user cannot be expected to have the time and patience to read through a large set of titles. Instead, the information access system should provide the user with tools to facilitate the assimilation of the results. One possibility is to help the user reformulate the query by suggesting alternative terms. Another possibility, explored in this paper, is to provide tools to aid the user in the navigation of the retrieval results.

Retrieval result navigation differs from the navigation of a document collection as a whole, an activity which is often referred to as *browsing*. An entire collection can be characterized by a fixed structure and so its contents can be organized into some *a priori* structure. By contrast, a user-specified search defines a particular slice through the contents of a collection. Even if predefined structure is available for the collection in its entirety, in many cases this structure is not appropriate for an *ad hoc* subset.

In this paper we show how a document clustering algorithm can be used to organize the results of retrieval into semantically-related document groups. The user can use these groups in several ways: to identify useful subsets of documents to be perused with other tools, to eliminate subsets whose contents appear nonrelevant, or to select promising document subsets for reclustering into more refined groups. This technique, which we call Scatter/Gather (Cutting *et al.* 1992), (Cutting, Karger, & Pedersen 1993), can expose the structure that is inherent in a document subcollection, in order to aid navigation of retrieval results.

Our experience with this use of clustering is that it often, although not always, produces groupings whose semantics can be inferred to a large extent by the user and used in the ways suggested above. Although the descriptions in this paper are anecdotal, we are in the process of obtaining quantitative evidence of the usefulness of the cluster information for organizing retrieval results, based on queries and relevance judgements from the TREC/TIPSTER collection (Harman 1993). We are also conducting user studies of the role of this tool in the context of an information access system that integrates Scatter/Gather with other tools.

In the remainder of this paper we describe the Scatter/Gather algorithm and illustrate its application to retrieval results via two examples. The first example makes use of encyclopedia text and a very general query in order to illustrate the kinds of groups that are produced by the clustering algorithm. Encyclopedia text is useful for such a demonstration because the main topics of the articles are relatively easy to infer from their titles. The second example consists of a more specific query on a more eclectic collection.

Scatter/Gather for Collection Browsing

Scatter/Gather uses the metaphor of a dynamic table-of-contents to help the user navigate a large collection of documents. Initially the system uses *document clustering* to automatically *scatter* the collection into a small number of coherent document groups, and presents short summaries of the groups to the user.

Based on these summaries, the user selects one or more of the groups for further study. The selected groups are *gathered*, or unioned, together to form a subcollection. The system then reapplies clustering to scatter the new subcollection into a new set of document groups, and these in turn are presented to the user. With each successive iteration the groups become smaller, and therefore more detailed.

The document clustering algorithm is optimized for speed, to encourage interaction, rather than to guarantee accuracy. The current system (see Figure 1) uses a linear-time clustering algorithm for *ad hoc* document collections and a constant-time algorithm for stable, preprocessed collections. The linear-time algorithm can organize 5000 short documents in under one minute on a SPARC20 workstation.

The cluster summaries are designed to impart general topical information. Clusters are summarized by presenting their size, a set of *topical terms*, and a set of *typical titles*. The topical terms are extracted from the *document profiles*, or weighted bag-of-words representations, of the documents included in the cluster and are intended to reflect the terms of greatest importance in that cluster. The typical titles are the titles of documents closest to the cluster centroid.

In our earlier work (Cutting *et al.* 1992), (Cutting, Karger, & Pedersen 1993), Scatter/Gather was used to organize the contents of an entire collection in order to allow the user to become familiar with its contents via query-free browsing. In this paper we begin to explore the application of Scatter/Gather to the navigation of retrieval results.

Scatter/Gather on retrieval results creates a kind of semantic characterization that is specific to the contents of the retrieved documents. This differs from a pre-specified hierarchy or set of category labels, defined independently of any query (or collection). As will be illustrated in sections that follow, the semantics of the groupings that result from a query's retrieval results may look idiosyncratic when compared against a general classification, but can be quite descriptive of the context of the query and the retrieved documents.

A General Query on General Text

In this section we illustrate the use of Scatter/Gather to help navigate an *ad hoc* collection. In this example the user is presented with a general text collection, Grolier's Encyclopedia (Grolier 1990), which contains articles relating to science, art, literature, history, and so on. To demonstrate the behavior of the algorithm, we assume the user issues a very naive query, a Boolean search on the highly ambiguous word *star*. The search returns about 400 documents, each of which much must bear some relationship to the concepts denoted by the word *star*, although as will be seen, that relationship has several dramatically different forms.

Initial Clustering

The documents are clustered into 10 groups (the number of groups is currently chosen arbitrarily but should be a function of the number of documents retrieved and the dispersion of the clusters generated), as shown in Table 1. We discuss each in turn.

Topical Terms	**Typical Titles**
Cluster 0 (6) ballet, sentence, language, balanchine, speech, dancer	Balanchine, George semantics figures of speech
Cluster 1 (7) station, broadcast, network, cut, program, gem, tv, cable	diamond radio and television Capital Cities/ABC
Cluster 2 (13) flag, rug, design, weave, carpet, pattern, stripe, force	medals & decorations flag rugs and carpets
Cluster 3 (21) game, player, team, league, ball, professional, football	Musial, Stan football soccer
Cluster 4 (65) energy, hydrogen, radiation, planet, temperature, gas	solar system astron. & astrophys. star
Cluster 5 (164) bright, stellar, constellation, magnitude, celestial	magnitude Barnard's Star multiple star
Cluster 6 (76) film, comedy, musical, singer, broadway, movie, hollywood	Murphy, Eddie film, history of Stanwyck, Barbara
Cluster 7 (21) theatre, stanislavsky, young, playwright, opus, debut	Bernhardt, Sarah acting O'Casey, Sean
Cluster 8 (16) novel, story, hemingway, key, book, writer, fiction, song	Bester, Alfred Hemingway, Ernest Star-Spangled Banner
Cluster 9 (11) sea, flower, sand, family, shape, specie, wind	brittle star echinoderm feather star

Table 1: Contents of cluster summaries for the results of a Boolean search on *star*. The number of documents in each cluster is shown in parentheses.

The order of the clusters is meant to impart a semantic continuum; for example, in our example the two clusters that pertain to the astronomical sense of star are contiguous (Clusters 4 and 5). Since the system uses a fixed number of clusters, there are usually one or two "junk" clusters that contain the documents

Figure 1: A portion of the top level of the Scatter/Gather interface over the Grolier's Encyclopedia.

that do not group well with the rest of the subcollection. These junk clusters exhibit the following properties: they have fewer documents than the other clusters, they appear at either end of the semantic continuum, and they are less coherent in topic than the other clusters.

In this example, Clusters 0 and 1 display these characteristics, as Cluster 0 contains six documents covering the two themes of ballet and semantics, and Cluster 1 contains documents describing broadcasting and gems. The use of *star* in these documents is not central to their meaning. Instead, it is used to indicate fame for a dancer and as an illustrative example in semantics (the morning star/evening star dichotomy). The clustering algorithm finds what is in common among documents; the use of *star* in these cases is only one component in that decision.

Cluster 2 contains documents with a military theme, in some cases describing famous military personal (*Oliver North, Chester Nimitz*) and in other cases military symbols (*medals and decorations, Uncle Sam, flag*). There are four "outliers" as well: one article discusses an entertainer and three articles relate to an astronomical sense of star.

Cluster 3 contains 21 documents relating exclusively to sports and games, a close cousin to the preceding military theme:

Contents of Cluster 3

Clemente, Roberto	Davis, Glenn
bowling	Ford, Whitey
Hagen, Walter	Carew, Rod
soccer	Nurmi, Paavo
Casper, Billy	Tilden, Bill
Luckman, Sidney	Robertson, Oscar
Johnson, Rafer	Beliveau, Jean
Chinese checkers	Best, George
Musial, Stan	Bench, Johnny
Rudolph, Wilma	football
Yastrzemski, Carl	

Cluster 4, with 65 members, and Cluster 5, the largest group with 164 members, both relate to astronomy, a topic which reflects a central use of *star*. Note that the topical terms in Cluster 4 are more focused on astrophysics, and this cluster is indeed the one that contains the articles that cover astrophysics. Thus the interface suggests the existence of a differentiation between astrophysics and astronomy in general, despite the fact that Cluster 4 contains documents of both types. Note also that in these clusters, articles about astronomers are intermixed with articles about astronomical phenomena. Samples are shown below.

Partial Contents of Cluster 4

light-year	nebula
extinction (in geology)	star
Oort, Jan Hendrik	stellar spectrum
three-body problem	Jeans, Sir James
solar system	life, extraterrestrial
Neptune (planet)	photosphere
planets	Local Group of galaxies
fusion energy	interstellar reddening
space exploration	brown dwarf
radio astronomy	astronomy, history of
relativity	proton-proton reaction
day	eclipse
Saha, Meghnad N	interstellar matter
angular momentum	ultraviolet astronomy
Digges, Thomas	gravitational collapse

Partial Contents of Cluster 5

black dwarf	Trumpler, Robert J.
Hoyle, Sir Fred	neutron star
Gill, Sir David	Argelander, Friedrich
Wolf, Max	magnitude
five-spice powder	subgiant
Milne, Edward Arthur	supernova
parallax	Libra
McDonald Observatory	coordinate systems
Payne-Gaposchkin, Cecilia	Pleiades
Star of David	Foucault pendulum
Agena	Bonner Durchmusterung
photometry, astronomical	Hyades
blink microscope	Olbers's paradox
Nikisch, Arthur	Star Chamber
Bjornvig, Thorkild	distance, astronomical

Cluster 5 also contains some "outlier" documents (all very short, thus making them more likely to cause errors) such as *five-spice powder*, (which contains *star anise*), *Star of David*, and *Thorkild Bjornvig* (a Danish poet), which are separated from the rest in the subsequent reclustering.

Cluster 6 consists of movie and film stars, Cluster 7 primarily of stars of theatre, and Cluster 8 of authors and written works that contain the word *star* in their title. Finally, Cluster 9 contains articles that describe animals and plants that have star-like structure or aspect.

Partial Contents of Cluster 8	Contents of Cluster 9
Caine Mutiny, The	feather star
McGrory, Mary	cowslip
McHenry, James	earthstar
Fort McHenry	echinoderm
Key, Francis Scott	blazing star
Cronin, A. J.	eelpout
nursery rhymes	star-of-Bethlehem
Williamson, Jack	brittle star
Oh, Sadaharu	shooting star
Star-Spangled Banner, The	sand dune
Cartland, Barbara	bishop's-cap

In this illustration, the algorithm appears strikingly adept at grouping documents according to general themes. There are some outliers, but because every document containing one or more instances of the word *star* is retrieved and must be placed in one of ten clusters, the partition will not be perfectly neat. In most cases, those documents that do have a common theme are indeed grouped together.

Reclustering

When the Scatter/Gather operation is rerun over the two astronomically-oriented clusters (4 and 5) the new clustering provides additional detail (see Table 2).

Each of Clusters 1-8 contain documents that cohere well to a common theme; they can be glossed with the labels Elements, Galaxies, Constellations, (Individual) Stars, Particles, Planets, Astronomers, and Navigation. The first and last clusters are the "junk" classes, and effectively separate out documents that clearly do not contain the astronomical sense of *star*. Note that in this reclustering the biographical articles are separated out from the rest and placed in their own cluster (Cluster 7).

A More Specific Query

In this section we demonstrate the use of Scatter/Gather on text whose content is more difficult to quickly assimilate: the TIPSTER collection of over 1 million newswire, newspaper, magazine and government articles, dating mainly from the late 1980's. We also make use of one of the TREC queries and its associated relevance judgments (Harman 1993). For this query, the task is to find all documents that discuss the following (abbreviated) topic:

Topic 87: Criminal Actions Against Officers of Failed Financial Institutions

We formulated a query containing the terms *bank financial institution failed criminal officer indictment*[1]

[1]Search terms that augment the original query state-

Topical Terms	Typical Titles
Cluster 0 (5) animal, conductor, nikisch, music, hypnosis, mesozoans	mesozoan Crumb, George Nikisch, Arthur
Cluster 1 (4) element, metal, atomic, weight, chemical, mirror,	telescope element hydrogen
Cluster 2 (8) galactic, spiral, globular, million, milky, hubble	Local Group of galax. extragalactic systems star
Cluster 3 (43) constellation, sky, northern, hemisphere, locate, north	Cygnus constellation Big Dipper
Cluster 4 (53) variable, sequence, spectral, white, diagram, pulsar	multiple star variable star stellar evolution
Cluster 5 (30) radiation, particle, emission, emit, region, source	radio astronomy star interstellar matter
Cluster 6 (37) planet, orbit, moon, jupiter, planetary, uranus, spacecraft	astronomy, history of solar system astron. & astrophys.
Cluster 7 (36) university, graduate, director, steven, dick, state, american	Schwarzschild, Karl Eddington, Sir Arthur Struve, Otto
Cluster 8 (5) altitude, navigation, ship, longitude, astrolabe	astrolabe, prismatic navigation astrolabe
Cluster 9 (8) david, curve, triangle, side, self, add, camargo	Star Chamber fractals Star of David

Table 2: Contents of Cluster Summaries after Scatter/Gathering on Clusters 4 and 5 from Figure 1.

and instructed the system to retrieve the 500 top-ranked documents according to a SMART-like weighting (Buckley, Allan, & Salton 1994). Out of these 500 retrieved documents, only 21 had been judged relevant to the query by the TREC judges (some may not have been judged at all, but for the purposes of this example, those with no judgement are simply considered to be not relevant). These documents were not ranked especially highly by the similarity search measure: none of the documents judged relevant appeared in the top 10, only one appeared in the top 20, and only four appeared in the top 40.

ment can be obtained by specifying the initial query and looking at the topical terms for the clusters obtained there. For example, one of the main kinds of financial institutions that failed during the time covered by the collection were savings and loans, *s&l*'s, and this term occurs as a topical term for Cluster 3.

A tool that can guide the user towards the relevant subgroups would indeed be useful. Here we show that the Scatter/Gather tool can be effective in this way. The system is instructed to gather the 500 documents into five clusters; below are shown the resulting sizes and topical terms:

Clusters and Topical Terms
Cluster 0 (4) assistant director deputy secretary special affair division administrator management staff position chief
Cluster 1 (187) deposit capital asset insurance risk fail save credit rate market account billion
Cluster 2 (217) section information 2 requirement regulation 3 request rule record 5 provision procedure
Cluster 3 (85) investigation allege fraud court lawyer firm prosecutor jury bcci american grand defendant
Cluster 4 (7) marcos philippine marcoses unite order export respondent racketeering khashoggi buy manhattan wife

Cluster 3 stands out for the purposes of the query in that it contains terms pertaining to fraud, investigation, lawyers, and courts. Note that in a general corpus these terms might not be descriptive for this query since the user would assume the documents were about legal issues in general. However, since we know the system has retrieved documents that also pertain to financial institutions, we can assume that the legal terms occur in the context of financial documents.

The topical terms for Cluster 4 are less compelling, for they have only one term corresponding to a crime and seem to clearly indicate documents discussing the scandal involving the leader of the Phillipines in the late 80's. The topical terms of Cluster 0 are very general (and there are only four documents in the cluster which can be quickly scanned). It appears that Cluster 0 contains very general documents that do not fit into any of the other clusters particularly well, whereas Cluster 4 contains documents that relate to a very specific allegation of fraud.

Cluster 1 is also compelling in that its summary contains many financial terms; however, it is less promising than Cluster 3 in that it seems more related to assets and risk assessment than criminal charges and failed banks. Finally, Cluster 2 seems related most strongly to rules and regulations, rather than indictments and fraud. Note again that this cluster, if taken out of context, might seem to refer to government regulations in general; however, since it was generated as the results of a query on financial terms, like Cluster 3, it most likely contains documents discussing rules and regulations on financial matters. A rescattering of the cluster confirmed this suspicion.

Based on this assessment, Cluster 3 looks most promising. If the user rescatters it, five new clusters are produced:

Clusters from Cluster 3 Above
Cluster 3.0 (4) sentence prison juvenile rough levine public father national liability alamo judge life
Cluster 3.1 (16) miami international manager credit branch customer allegedly country loan noriega cocaine authority
Cluster 3.2 (29) security tax sec trade convict seek trial judge stock kidder house money
Cluster 3.3 (28) loan thrift save texas real estate s&l fail dallas guilty plead 1985
Cluster 3.4 (8) american clifford safra altman executive washington cantor peru man greco latin article

The user might choose the first, third and fourth clusters since their topical terms all seem to pertain to the topic of interest. Clusters 3.2 and 3.3 are especially compelling since they contains terms pertaining both to finance and to criminal proceedings. Cluster 3.2 has more terms about conviction but 3.3 has more terms pertaining to failure and the kinds of financial institutions that the user may have known to have failed; namely S&L's and thrifts. As it turns out, the clusters' contents reflect these observations: Cluster 3.2 contains mainly articles about indictments pertaining to financial fraud involving securities and stocks, but not failed banks.

The user can view the contents of a cluster in ranked order, according to the score generated by the similarity search, or can view the documents according to some other search tool. Based on the topical terms, the most promising looking clusters are 3.0, 3.2, and 3.3. It turns out that Cluster 3.0 has one relevant document out of four:

Law – Legal Beat: U.S. Court Backs Effort by Insurers to Avoid Liability in S&L Failures

The other documents in this group are:

Ex-Director Of NY Fed Sentenced To Six Months For Data Leak
Thankless Task: Tough Young Clients Generate Tough Cases For A Public Defender
Law: Bank Mergers May Be Complicated By Issues Of Liability In Alamo Case

The third cluster, Cluster 3.2, has one relevant document out of 28:

Federal Jury Indicts Keating And Associates – SEC Also Files Civil Charges Alleging Insider Trading

High ranked documents in this group include indictments for other crimes, some of which are financial in nature:

Law – Legal Beat: Illinois Judge Acquits Executives In Major Workplace Injury Case
Poindexter, North, Secord, Hakim Indicted By Iran-Contra Jury
Gaf Is Indicted In Alleged Scheme To Manipulate Carbide Stock In '86
Drexel Learns U.S. May Soon Ask An Indictment From Grand Jurors
Broker, Ex-Chief Of Bank Are Charged In New England Insider-Trading Inquiry

Cluster 3.3 has eleven relevant documents out of 29. Some of these are:

Four Plead Guilty In U.S. Inquiry Of Bank Violations
U.S. Spotlights Suspected Bank Fraud In Texas But Skeptics Ask If Broad Effort Is Fated To Fail
Five Named In Indictments Stemming From Failure Of Florida's Sunrise S&L Ex-Ownder Of Texas Thrift
Ex-Chief Of Failed Empire S&L In Texas, Six Others Indicted On Fraud Charges

The second cluster also contains two relevant documents. Most other documents in this cluster discuss indictments for money laundering along with one article involving Noriega and another on a teen scandal in San Francisco. The last cluster has no relevant documents although it has several that discuss the BCCI.

So it turns out that Cluster 3 contains 15 of the 21 relevant documents. The remaining 6 relevant documents are found exclusively in Cluster 2. When this cluster, which contains 187 documents, is scattered, all six relevant documents appear in one cluster of size 107 whose key terms are: *insurance save thrift fdic texas depositor regulator s&ls insure chairman real estate*.

Related Work

Space limitations prevent a detailed exposition of related work, so only a few approaches are touched on here. Several systems display document collections in what can be described as a similarity network. A focus document, usually one that the user has expressed interest in, is shown as a node in the center of the display, and documents that are similar to the focus document are represented as nodes linked by edges surrounding the focus document node. Systems of this kind include the Bead system (Chalmers & Chitson 1992) and the system of (Fowler, Fowler, & Wilson 1991).

Another common approach is to define a concept hierarchy or lattice of terms or attributes that represents the entire document collection, and have the user navigate this lattice. Another version of this approach is to allow the user to search only those attributes or terms that the user has specified. Systems using one or both of these approaches include those of (Korfhage 1991), (Arents & Bogaerts 1993), and (Carpineto & Romano 1994). The Cougar system (Hearst 1994) is tailored more towards retrieval results, and displays all and only the attributes that are actually assigned to retrieved documents. It has an advantage over Scatter/Gather and most other category-assignment systems in that it allows documents to be placed in more than one category simultaneoulsy. However, it relies on pre-defined categories which are not always appropriate for the *ad hoc* nature of retrieval results.

Summary

We have demonstrated that a clustering algorithm can be used to induce structure from a collection of documents to aid navigation, making use only of content similarity among the documents. This point is important because sophisticated analyses are typically not available for an *ad hoc* collection. For example, hand-assigned topic labels might not be as useful as desired for a set of retrieval results, because the query might have retrieved documents most of which are assigned the same label. Scatter/Gather essentially uses the collection as a knowledge base to present a topic-coherent structure for browsing. It remains to be seen whether algorithms that are "smarter" about the topical information they assign to the documents are more useful for such a task.

As mentioned above, we are in the process of evaluating this use of Scatter/Gather in the context of a more complete system. Notably, this system allows the user to display the contents of one or more clusters in terms of TileBars (Hearst 1995), a graphical display paradigm that can simultaneously and compactly indicate the relative length of the document, the frequency of the query terms in the document, and the distribution of the query terms with respect to the document and to each other. Thus the Scatter/Gather clusters allow the user to select a promising subset of the documents, and the TileBars display, when used on this smaller subset, facilitate assessment of the relevance of the documents to the query.

References

Arents, H. C., and Bogaerts, W. F. L. 1993. Concept-based retrieval of hypermedia information – from term indexing to semantic hyperindexing. *Information Processing and Management* 29(3):373–386.

Buckley, C.; Allan, J.; and Salton, G. 1994. Automatic routing and ad-hoc retrieval using SMART: TREC 2. In Harman, D., ed., *Proceedings of the Second Text Retrieval Conference TREC-2*. National Institute of Standards and Technology Special Publication 500-215.

Carpineto, C., and Romano, G. 1994. Dynamically bounding browsable retrieval spaces: an application to galois lattices. In *Proceedings of RIAO '94; Intelligent Multimedia Information Retrieval Systems and Management*, 533–547.

Chalmers, M., and Chitson, P. 1992. Bead: Exploration in information visualization. In *Proceedings of the 15th Annual International ACM/SIGIR Conference*, 330–337.

Cutting, D. R.; Pedersen, J. O.; Karger, D.; and Tukey, J. W. 1992. Scatter/Gather: A cluster-based approach to browsing large document collections. In *Proceedings of the 15th Annual International ACM/SIGIR Conference*, 318–329.

Cutting, D. R.; Karger, D.; and Pedersen, J. 1993. Constant interaction-time Scatter/Gather browsing of very large document collections. In *Proceedings of the 16th Annual International ACM/SIGIR Conference*, 126–135.

Fowler, R. H.; Fowler, W. A. L.; and Wilson, B. A. 1991. Integrating query, thesaurus, and documents through a common visual representation. In *Proceedings of the 14th Annual International ACM/SIGIR Conference*, 142–151.

Grolier. 1990. *Academic American Encyclopedia*. Danbury, Connecticut: Grolier Electronic Publishing.

Harman, D. 1993. Overview of the first Text REtrieval Conference. In *Proceedings of the 16th Annual International ACM/SIGIR Conference*, 36–48.

Hearst, M. A. 1994. Using categories to provide context for full-text retrieval results. In *Proceedings of RIAO '94; Intelligent Multimedia Information Retrieval Systems and Management*, 115–130.

Hearst, M. A. 1995. Tilebars: Visualization of term distribution information in full text information access. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*. Denver, CO: ACM.

Korfhage, R. R. 1991. To see or not to see – is that the query? In *Proceedings of the 14th Annual International ACM/SIGIR Conference*, 134–141.