Process Evaluation in eDiscovery as Awareness of Alternatives

Jeremy Pickens, John Tredennick, Bruce Kiefer
Catalyst Repository Systems
1860 Blake Street, 7th Floor
Denver, Colorado
303.824.0900

{jpickens, jtredennick, bkiefer}@catalystsecure.com

ABSTRACT

With a growing willingness in the legal community to accept various forms of algorithmic augmentation of the eDiscovery process, better understanding of the quality of these machine-enhanced approaches is needed. Our view in this position paper is that one of the more important ways to understand quality is not in terms of absolute metrics on the algorithm, but in terms of an understanding of the effectiveness of the alternative choices a user could have made while interacting with the system. The user of an eDiscovery platform needs to know not only how well an information seeking process is running, but how well the alternatives to that process could have run.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – Search process

General Terms

Measurement, Experimentation, Standardization.

Keywords

Iterative Information Seeking, Interactive Information Seeking, eDiscovery, Process Evaluation

1. INTRODUCTION

Unlike traditional ad hoc search (such as web search) in which the information seeking process is typically single-shot, eDiscovery has both the potential and the necessity to be iterative. Information needs in eDiscovery-oriented information seeking are changing and ongoing, and often cannot be met in a single round of interaction or by a single query. Evaluation of eDiscovery platform quality must take this into account.

There are many metrics for single-shot, non-interactive retrieval, such as precision, recall, mean average precision, and PRES [1]. Our goal is not to propose a new single-shot metric. Instead, we declare that what is needed is an approach in which any or all of these metrics are used in an interactive context.

Furthermore, we take a user-centric view in that we are not concerned with comparison between eDiscovery platforms but instead are concerned with helping the user understand where he or she is within a larger information seeking task on a single platform. A quality process should be one in which the user is able to both (1) affect system behavior by making conscious choices, and (2) explicitly obtain an understanding of the consequences of those choices, so as to adapt and make better choices in the future.

2. THE "WHAT IF" OF EDISCOVERY

2.1 Choices

Interactive information seeking in general and eDiscovery in particular are characterized by choices. Even with machine augmentation of the search process there is still a human in the loop, considering alternatives and making decisions. Examples of choices, not all of which are independent of each other, include:

1. Does one continue traversing the results list for an existing query, or does one issue a new query instead

2. If complete queries are offered as suggestions, which of the alternatives does one pick?

3. If individual terms are offered as query expansion options, which of the alternatives does one pick, and when does one stop adding additional terms?

4. If the collection is clustered in some manner, which cluster does one choose to examine, and when does one stop examining that cluster?

5. If multiple sources (e.g. custodians) or document types (e.g. PDF, PPT, Word, email) are available, how does one choose which sources or types to pay the most attention to?

6. When the document volumes go beyond what is feasible to review, how do you determine when to stop reviewing?

7. At what point do you produce documents which haven't been personally reviewed.

2.2 Consequences

In the previous section we outlined a few examples of the types of choices that an information seeker has to make. Each of those choices has consequences. The choice to dedicate time and resources investigating information coming from one custodian means that less time and fewer resources will be dedicated to a different custodian. More time spent traversing the result set of one query means less time spent on the results of a different query, or perhaps fewer queries executed overall. Adding some terms to an existing query (during query expansion) means not adding others. Deciding that a particular point would be a good one at which to stop reviewing, and then continuing to review anyway might yield diverging expectations as new pockets or rich veins of information are discovered.

In order to understand the quality of a search process, knowing the effectiveness of such choices are not enough. A user has to be able to come to know and understand the opportunity costs of the choices not taken. Does an eDiscovery platform make it possible for a user to understand the consequences of his or her choices?

Does the system give a user a working awareness of the alternatives? Is it possible for the user to return to a previous choice at a later point in time and obtain feedback on the question of "what if" that path had been chosen? A quality search process should be able to answer, or at least give insight into, these questions.

3. PRINCIPLES AND EXAMPLES

Giving an information seeker an awareness of alternatives is not an approach tied to any one particular algorithmically-enhanced methodology. The manner in which a machine (algorithm) learns from the human and applies that learning to the improvement of future choices is a separate issue from whether or not the user is able to garner insight into the efficacy of alternative choices. Granted, some algorithmic approaches might be more penetrable, more conducive to proffering the needed awareness. But the feedback on choices taken versus not taken are going to depend heavily on the nature of the choices themselves.

That said, we offer a few principles which might aide in the design of consequence-aware systems:

1. If there is overlap between the multiple choices (i.e. if the consequences of certain choices are not mutually exclusive) then information garnered while following one choice could be used to make inferences about another choice.
2. If there is overlap between the consequences (results) of a single choice, then the efficacy of that choice can be more quickly assessed by examining fewer, perhaps more "canonical" results.

For example, a clustering algorithm might not partition a set of documents, but instead place a few of the same documents in multiple clusters. Or the same (duplicate or near-duplicate) documents might be found in the collections from more than one custodians. Or two different query expansion term choices (e.g. "bees" and "apiary") might retrieve many of the same documents. In such cases, judgments (coding) on these shared documents can be used to assess multiple choices. Naturally the assessment is done within the context of whatever metric is most important to the user, whether that metric is precision, recall, or something else entirely. But the principle of using overlap to estimate and make inferences on that metric remains.

The way in which this could be made to work would be to implement a process-monitoring subsystem that keeps track of choices both taken and not taken, and then uses information such as the ongoing manual coding of responsiveness and privilege to assess the validity of those choices. The differential between expectation at one point in time and reality at a future point in time should yield more insight into the information seeking eDiscovery process than just knowing the precision or recall effectiveness at any given point in time.

4. ISSUES

The largest issue that needs to be resolved for alternative-aware approaches is that of ever-expanding choice. At every round of interaction, at every point in the human-machine information seeking loop at which the human has the ability to make a choice, a number of options become available. Every choice then gives rise to another set of choices, in an exponentially-branching set of alternatives. Naturally this exponential set needs to be pruned into a manageable set of the most realistic, or possibly the most diverse, set of alternatives.

The consequences of every possible choice or path not taken probably do not to be tracked and monitored; a subset should be fine. However, there needs to be enough awareness of alternatives that the user can get an overall sense of how well he or she is doing, and how much progress is or is not being made with respect to the other choices that were available at various stages. The user needs to be able to get a sense of how well a choice at one point in time matches reality as the consequences of that and other, hypothetically-followed choices become clearer at later points in time.

5. SUMMARY

Information retrieval has a long history of using user interaction (e.g. in the form of relevance feedback and query expansion, for example) to improve the information seeking process in an iterative manner. User behavior alters the algorithm. However, it is also true that the algorithm alters the user. The more choices a user makes, the more potential exists that some of these choices are sub-optimal. Therefore, awareness of alternative choices are needed to help the user orient himself inside of complex information seeking tasks such as in eDiscovery. This paper proposes an approach to the evaluation of quality not in terms of system comparison, but in terms of alternative, path-not-taken choice comparison and awareness.

6. REFERENCES

[1] Magdy, Walid and Jones, Gareth. *In the Proceedings of the 33rd Annual SIGIR Conference.* PRES: A Score Metric for Evaluating Recall-Oriented Information Retrieval Applications. Geneva, Switzerland. August 2010.