Odin: Collective Intelligence on the Go

Joshua Hailpern
HP Labs
1501 Page Mill Road Palo Alto, CA 94304
joshua.hailpern@hp.com

Bernardo A. Huberman
HP Labs
1501 Page Mill Road Palo Alto, CA 94304
bernardo.huberman@hp.com

ABSTRACT

A daunting consequence of the prevalence of the web and user-generated content is that information, which used to be scarce and therefore valuable, is now plentiful. Consequently, what has now become scarce, and therefore valuable, is the user’s time and attention. Furthermore, with the emergent ubiquity of mobile devices, users are expected to be up-to-date on the latest reports, appraise financial outlooks, and be ready to have an informed discussion no matter where they are. Solutions must be designed to meet daily needs of users of mobile devices, requiring methods for finding what to read, and ways of presenting it within the small visual real estate for users with limited time and attention. To this end we present Odin, a mobile web-based window onto a user’s document corpus. Instead of corpus summarization, Odin uses machine learning algorithms to examine a corpus of documents, determines what the documents discuss, and how the crowd is talking about them. Based on this holistic view of the crowd’s wisdom, Odin then ranks each document and opinion within each document and relates them to the crowd’s overall impression. This information is presented through a simple and intuitive interface. With Odin, mobile device users can quickly find key opinions that are very Relevant, highly Aligned with, or Divergent from the crowd’s point of view.

Keywords

Wisdom of the Crowd; Mobile; Interface; Interaction; Economics of Attention; Alignment

Categories and Subject Descriptors

H.5.m. [Information Interfaces and Presentation (e.g. HCI)]: Miscellaneous

General Terms

separate the terms with a semi-colon

1. INTRODUCTION

While the web enables information access on a scale that a few years ago would have been inconceivable, it also creates the problem of information overload; access to content that is relevant in a timely manner, without distractions from information which is marginal in value to the user. The problem becomes even more acute when we acknowledge that people are increasingly using their mobile devices for their daily needs. Equally requiring methods for finding what to read, and ways of optimizing this content for mobile devices’ small visual real estate. Users must be able to securely relevant information quickly, with little overhead. While this need has permitted many document-centric knowledge workers (e.g., political analysts, financial investors, economists, CEOs), legal professionals (e.g., clerks, layers, judges) and other consumers of topical information (e.g., reviews, gossip blogs), a mobile device user’s needs are dependent upon how much time that user has to read documents. For the executive that has a matter of minutes before a meeting starts, she only has time to read one document, and maybe not even its full length. However, the lawyer waiting for about a fifteen or twenty minutes for his partners to arrive, can perhaps read two or three long documents. Meanwhile the financial analyst making an hour long train commute to work, may have more leisure to fully browse a corpus, though her time and UI must still be optimized as much as possible. Though each of these temporal and interaction constraints are different, the underlying goals of the user are constant: on a mobile device, quickly traverse a corpus of documents, find the number of documents that meet a time constraint, and naturally find key points in each document based upon a broader context.

In response, we created a novel interaction model and system called Odin. This work aims at exploring our understanding of attention allocation [45, 24, 25, 1] so that mobile device users within and beyond enterprise can browse and find the most salient documents in a large corpus, and quickly find what critical opinions are within each document based on the wisdom of the crowd. This Odin system is tailored for mobile devices and users that have limited time to consume documents. Unlike corpus summarization, Odin’s analysis algorithm leverages machine learning algorithms to examine a corpus of documents, determines what topics the documents discuss, and how the crowd is talking about them, to help users zero-in on the most salient documents and opinions by placing and ranking each opinion within the context of the crowd.

Odin itself employs three key designs to meet a users’ varying needs. First, to help find a document, Odin ranks each document on three different scales: how Relevant, Aligned and Divergent it is from other documents in the crowd/corpus. Thus, a user can quickly find exactly the document that they are looking for. Second, regardless of how a user finds their way to a document, Odin provides an Executive Summary of each document, highlighting key sentences and opinions for easy document digestion. As a result users need not read the full text (though it is still easily accessible), as it can often be long and challenging to read on small screens. Rather users can focus on only the most important pieces of information. Third, the Odin UI has quick one-tap interactions to allow
a user to find the most Relevant, Aligned or Divergent document without browsing. This is a critical design consideration for users that have a very limited time and must get in-and-out of a document corpus as quickly as possible.

The foremost contribution of this work is a system, an algorithm and a UI interaction technique that allows the user to quickly find the document(s) needed given time constraints, focus on the most important opinions in any document, while still being able to see how each article and opinion fits in its context. We first discuss how Odin builds on the existing literature and is situated within the broader set of solutions. We then present Odin’s features and implementation, followed by positive results from an initial evaluation we conducted, and conclude with a discussion of future ongoing work.

2. RELATED WORK

Issues pertaining to information overload are not new, and we codified in Simon’s description of the relationship between attention and information [45]. While research across many disciplines has expanded on the notion of an economics of attention [16, 26], information in the digital age has been shown to be a function of user access [1, 24] and presentation [25]. Thus, we can look at the findings on economies of attention as a strong grounding for interfaces that increase the speed and access to large document corpuses and the key opinions within them.

It is important to note that while much work in the machine learning community has focused on automatic document summarization [5, 33, 47], Odin is not a document summarization system or algorithm. Odin focuses on user interaction design, and systems to support the navigation of opinions within a corpus on mobile devices by placing opinions and documents the context of a corpus.

2.1 Reading Documents

Researchers on document consumption has identified four key activities: reading, annotating, collaborating, and authoring [10]. While all four are important, Odin focuses on the first activity reading, as it pertains to exploring the overwhelming collection of documents in a corpus.

Unlike user-generated reviews, which are generally short and easily consumed, reading documents is a more arduous task due to their increased length and complexity. When reading documents, readers enter a “planning phase” [35] when moving within and between documents. During this phase, readers focus on finding important facts so as to connect key “bits” of information. For mobile users, the ability to quickly find key information is even more important [13]. Cui suggests that because mobile devices do not allow users to users to see multiple documents at the same time, people must keep all salient information in working memory. Therefore any solution should directly support the planning phase of reading, as well as minimizing cognitive overload in finding the “right” document and key opinions because a user’s working memory is in such active use.

2.2 Guiding Users During Reading

Technological solutions are well situated to help users wade through a large document, and find key sentences on which to focus. AKTiveMedia [12], Fishnet [8], and others [43, 46, 9] use vibrant color highlighting within a document to draw users’ attention to keywords based on queries issued. These techniques have proven successful in a desktop metaphor, and we build upon their success by using color to highlight key sentences in a long document. In a similar vein, Ahmed’s recent work on accessible skimming [2] attempts to merge document summarization and screen-readers to help visually impaired users find key points within a document. Given this success for interactive reading tools, we believe that there is much evidence for this need, and room for improvement.

2.3 Contextual Reading

Beyond supporting the process of reading itself, many researchers have been examining document context. Given that documents no longer exist in a vacuum, solutions must be created that help situate a document in the broader context, or help users choose a document based on said context.

For example Souneil Park [41, 40, 37, 38, 39] has attempted to explore bias in news events by showing pairs of documents on contrasting ends of a political spectrum side-by-side. While this technique helps with directly comparing two known contrasting opinions (via two documents), it does so by showing contrasting documents side-by-side. Existing literature suggests that this UI decision makes it more difficult to traverse content [35], and clearly would not map onto a mobile platform.

At a corpus level, tools can help users organize a document space. Popcorn [14] and many other tools [21] employ a wide range of visualizations to help users traverse a collection based on hierarchical semantic relationships (e.g. tree maps, network graphs). However, many of these visualization styles can be confusing [21, 11]. For example, Tag Clouds are actively used in many systems (e.g. [30, 51, 18, 32, 31]) and are intended to provide a high level view of a document space. However, evidence suggests that while they may be aesthetically pleasing, from a content consumption perspective they are unusable, misrepresentative, and inferior to simple sorted lists [22, 30, 42, 20]. An alternative Tag Cloud visualizations are systems that utilize textual summaries/synopsis [33, 29] or key quote lists (e.g. Yelp or Amazon.com) to provide smoother, more natural means of conveying key critical information and assisting users in finding salient information. While text summaries are useful for generally understanding a corpus at a high level, they do not help users trace those summaries back to specific documents/opinions; nor do they allow users to traverse a document space itself. Further, while many of the above solutions help organize content by topic, they do not provide assistance to locate documents based on the relationship between a document’s opinions and those of the crowd (are they aligned or divergent). It is in the space of unfulfilled needs that we situate our work on Odin.

2.4 Reading on Mobile Devices

Unlike desktop devices, mobile device UI must be optimized both for the smaller screen-size and “on-the-go” use case of users. While not an explicit research solution, many document-based websites use pagination to render their content. Pagination breaks long document content up into small sections, each residing on a different webpage that must be manually traversed by a user. However, empirical studies have shown that users greatly prefer scrolling text/lists [44, 19] and can easily read informational text up to 20 screens in length [27]. Further, users have the fastest information extraction when navigation is broad, with the fewest levels before reaching the desired content [19]. Even when changes occur through AJAX loads (rather than loading new pages), users are challenged and have a difficult time seeing and mitigating change on mobile [44]. Based on this work, solutions must be created that get users to documents and content quickly, without pagination, while still providing context.

Traditionally, mobile UI research on documents has focused on issues of layout, converting multicolumn pages into mobile-friendly UI [50, 28, 52]. However, as smart phones have become more prevalent, new applications and problems have arisen. We there-
fore see a striking parallel between the above work on contextual reading, and the new breed of smart phone UI.

According to Huang [23], “none of that work has investigated opinion mining... for a smart phone.” Huang’s paper on RevMiner[23] provides a mobile UI system that helps users understand reviews on restraints. Their UI is optimized for the small screens of mobile devices by utilizing a multi-facet view on the review data. We build upon RevMine’s successful findings on reviews, while simultaneously moving into a more complex documents data space, which tend to be longer and more nuanced. In addition, we shy away from more abstract views, and focus on presenting key sentences with full context. In addition, we provide a ranking system which highlights opinions that are aligned with or contrary to the views of the crowd rather than just being positive or negative.

Otterbacher [36], in a similar vein to [2], created a hierarchical document summarization UI for non smart-phones. Unlike RevMiner, Otterbacher’s system targets documents. We directly build upon their success, which highlights the power of providing users key facts about a document as a quick method for accessing information from long text.

3. SCOPE & MOTIVATION

While existing solutions to solving problems related to economics of attention and document corpus overload have been desktop based, Odin examines this problem space in a mobile context. The additional constraints due to the device (e.g., small screen real-estate), and the context of use (e.g. users with greatly decreased time availability to browse and consume) greatly change the functioning of the resulting solutions.

In this manner, Odin focuses on finding and bringing crucial sentences and documents to the user’s attention. Odin automates this process by gathering the collective intelligence of the crowd (via a corpus of documents), distilling what “the crowd” is talking about, and examining how the crowd refers to those topics. Odin then presents the user with documents and key sentences within each document based on the calculated context/alignment to the crowd’s opinions. It is important to note that Odin’s system and algorithm does not perform summarization. Rather, Odin focuses on using opinion alignment to improve document selection, and opinion highlighting, to make for a more meaningful and wide selection of crucial opinions.

In this paper, we will refer to sentences whose opinions are in-line with those of the crowd as Aligned, and sentences with opinions that go against the grain of the crowd as Divergent. A weighting algorithm determines the rank, or presentation order, of each document and sentence based on how important its topics are, and how much the opinions are Aligned with or Divergent from the crowd. We also created a Relevance ranking which can be used to find the sentences or documents that are the most opinionated (regardless of if the opinions are Aligned or Divergent from the crowd). Relevant documents can therefore be thought of as documents that provide a broad perspective.

4. Odin

Odin is comprised of a web-based UI, that allows a user interact with documents and the broader corpus, and a back-end algorithm and system, that performs the data mining analysis of the document corpus. Based on the existing literature, Odin seeks to address five key design requirements:

- **R1 Support Planning Phase of Reading:** Help readers quickly find important facts with little cognitive effort [35] from a range of opinions from the perspective most applicable to a given user’s needs

- **R2 Support Finding Key Documents:** Reduce cognitive load from managing multiple documents on mobile devices [13] from a range of opinions from the perspective most applicable to a given user’s needs

- **R3 Mitigate Document Length:** Use visual techniques to help users find key sentences in documents regardless of document length [12, 8, 43, 46, 9]

- **R4 Document Selection by Alignment:** Allow users to find documents based on relation to crowd’s opinion (unlike existing solutions which focus on summary only)

- **R5 Context Based Interaction:** Rather than a high-level abstract UI (unlike summarization tools), allow users to see all text within an individual document’s context [36]

The Odin UI itself is rendered in the web-browser of any mobile device. Screenshots that appear in this paper were taken from an iPhone5. In the following section, we detail the UI interactions as perceived by the user. The section following the UI discussion covers the specifics of how the underlying data mining analysis makes its calculations and models. Odin has been tested on political and financial reports, as well as the more mundane domain of movie reviews.

5. DESIGN & INTERACTION

Odin is designed to allow users to quickly manage a flood of information, both in terms of document corpus size and length of
documents themselves. We therefore made multiple design decisions to optimize the screen release to minimize the time needed to access the key documents and opinions, so a user can easily access as much information as possible given time constraints. In this regard, while Odin’s algorithm may be able to calculate a large amount of numerical analytics about documents and opinions that could be presented through data visualization, we focused on presenting only the most critical content as possible to minimize any visual destruction or interaction complexity.

To illustrate the breadth of the features and functionality in Odin, we present a series of scenarios that follow Anthony, a financial analyst:

Anthony is a junior financial analyst for Kirby, Lee & Co. On October 19th, 2012 Donald heard about Google’s accidental earnings report release, and the negative Google’s financial outlook. This was shocking news, and Anthony must quickly get a grasp on the flood of financial reports coming in and get into a meeting with his boss to discuss how they should handle their company’s investments. Anthony loads a document corpus of financial articles related to the premature announcement into Odin.

We present three scenarios each where Anthony has progressively more time to react. The scenarios illustrate the novel UI interactions (quick select buttons, Executive Summary, and contextual highlighting), and perspectives into a corpus (Relevance, Alignment, and Divergence). Multiple overly extensive scenarios are needed to illustrate the multiple approaches and uses of Odin. In practice, users may apply any subset of the following techniques at any time or in any order.

5.1 Scenario 1: a Matter of Minutes

As the news comes across his desk about Google’s announcement, Anthony gets a worried email from his boss Mr. Borson “Get into my office NOW, we need to figure out what to do ASAP.” Antony grabs his mobile phone, and heads from his office to Borson’s. This is a walk that will take less than five minutes, but in this time, Anthony needs to get up to speed on the issues at hand.

Anthony opens up Odin with a dataset of financial articles about Google’s premature announcement, and is presented with the Odin splash screen (Figure 1). While Odin has hundreds of documents, Anthony only has time to read one. Whichever document Anthony reads must provide him with the most information possible (R2).

On the Odin splash screen, Anthony taps the large green button labeled “Most Relevant.” This loads the Executive Summary (Figure 2a) for the most Relevant document (based on a weighting algorithm detailed in Section 6), which is entitled “Google shares suspended after accidental email wipes ...” We can conceptually think of the “Most Relevant” document, as the document with the broadest spectrum of opinions; those that are both representative of the corpus and those that are out of line with the corpus. It is the document that allows Anthony to get up to speed the fastest.

Rather than displaying the full content of the most Relevant document (which can be quite long and arduous to read on a small screen), the Executive Summary finds the sentences in the document with opinions, ranks them based on how Relevant they are compared to the opinions of the crowd, and displays them in rank order.

---

1 This scenario in no way reflects opinions on the value or worth of Google stock. This is one of our test data-sets we used, and constructed a narrative around it.
order (R1 & R4). This order is often not in the same chronology as they appear in the original text, but rather, highlight those sentences and opinions which provide the most interesting information to the reader. Anthony can quickly and easily determine how Google and their finances sit, the company’s outlook, and how people are discussing the company. Therefore top opinions are easy to find regardless of the length of the original document or where they occur in the source text (R1 & R3).

As Anthony quickly skims the most Relevant sentences in this document, he finds one that stands out. In the Executive Summary UI, Anthony taps the sentence and Odin loads the full document text. Odin then highlights the sentence Anthony tapped in green similar to a highlighter (Figure 3a), then Odin auto scrolls the full text to the location of the highlight. Now Anthony can see the full context of that sentence (R5). As Anthony turns the corner to enter Mr. Borson’s office, he is as ready and prepared as he could be for this spontaneous meeting.

5.2 Scenario 2: Meeting in 15 Minutes

We begin our scenario again, this time Mr. Borson asks for an all hands meeting in ten minutes. Anthony has only a short time to get prepared and ready to discuss the Google news alert. Anthony grabs his phone, heads to the meeting room (he does not want to be late) and begins to use Odin to do research until the meeting begins.

Given a slightly more relaxed time constraint, Anthony wants to explicitly see what the crowd’s conscious is about the Google announcement. In addition, it is also quite useful to know what outlying opinions are being voiced. On the Odin splash screen Anthony directs his attention to the two other quick elect options: “Most Aligned” in gold and “Most Divergent” in red (R4). Similar to the “Most Relevant” button, these other two quick select options will load the “most” Aligned or Divergent document into the Executive Summary view (Figure 2b and 2c respectively).

Within the Aligned Executive Summary (Figure 2b), Anthony can see the general consensus for Google, which appears to focus on the company’s mobile outlook. In short, Google has not monetized their mobile platform well. Likewise, the Odin UI also allows Anthony to easily see the divergent opinions. Here, the power of Odin truly shines. Figure 2c is a screen shot of Anthony’s device, showing the most Divergent opinions in the most Divergent document.

The most Divergent opinion in the entire corpus is a quote, “Our business had a strong quarter”. When Anthony sees such a strikingly different point of view that clearly goes against the grain, Anthony wants to see more. As in Scenario 1, Anthony can easily view any sentence in its respective document’s context with a single tap, and clicks on the quote in question. He can now easily see that this is a quote from the chief business officer, who is attempting to spin lukewarm results in the best way possible. While this is so different from the consensus, Anthony now can see Google’s perspective, why Google thinks they should be viewed positively, and Anthony also now realizes that these views are not shared by the financial analysts.

5.3 Scenario 3: On the Train-Ride to Work

The final scenario begins early in the morning before Anthony gets to work. As Anthony is packing up for his morning commute, Mr. Borson sends him a message requesting an all hands meeting to discuss the premature Google announcement once when everyone is at work. As Anthony begins his commute to work on the train, he has roughly a thirty minute window of time to get caught up with the financial reports to be discussed. Because the train is not the place to pull out a laptop, he turns on his phone and uses Odin.

As in the above scenarios, Anthony quickly skims the three quick selection documents, getting a broad overview of key opinions with the most Relevant Aligned, and Divergent documents. These provide him an overview of the situation. However, given the increased amount of time available, Anthony wants to delve deeper and expand his understanding of the issue, and the opinions of the financial analysts.

Running along the bottom of the Odin splash screen are three smaller colored buttons (Figure 1). Each button when clicked, loads the full corpus of documents sorted by how Relevant, Aligned, or Divergent the document is (Figure 3b). The color of the button corresponds with the sort order (green for Relevant, gold for Aligned, and red for Divergent). Upon tapping one of the full document list buttons, a list of each document (R2) sorted by Relevance, Alignment, or Divergence. Anthony can see the sort order selected (along the top of the UI). Each document in the list is presented with its title, and a preview of the first sentence in the document. Anthony can now select any document to see the Executive Summary for that document. The Executive Summary of key sentences is sorted in the same manner as the document sort list (e.g. if the document sort was by how Aligned the documents were, the sentences in the Executive Summary would likewise be sorted by how Aligned each sentence was).

As Anthony beings traversing documents and their corresponding Executive Summary, he finds one document to be very compelling. Rather than looking up each key sentence’s context individually, using the tap and highlight interaction, he decides to read the document in full. Along the bottom of the Executive Summary UI, Anthony taps the words full text. This loads the full text of the document much like Figure 3a, though without the colored highlight. Anthony can now read the document in full. As Anthony gets of the train, and enters his office he has perused dozens of documents, getting a quick breadth of opinions, while also being able to gather a depth of understanding. Thus he is able to discuss where there is commonality and diverging opinions in the crowd and on the issue of Google’s premature stock announcement.
6. IMPLEMENTATION

The Odin interface is built in HTML, CSS and JavaScript. Document corpora undergo a Java-based six-stage algorithm in order to generate the content for the Odin UI (Figure 2 & 3): 1) Content Extraction; 2) Identification of Discussion Subjects; 3) Detect Opinion Sentiment; 4) Calculate Crowd’s Opinions; 5) Sentence Level; and, 6) Ranking and Weighting. This algorithm provides the analysis to directly support R4, which facilitates the UI components that satisfy R1 and R2.

It should be noted that the Odin assumes that a corpus of documents (PDFs or links to webpages) on the same subject have already been collected.2

6.1 Stage 1. Extract Content

Regardless of document format (e.g. PDF, HTML, DOCX, etc.), the raw paragraphs, sentences, and words must be extracted in order to perform any ML analysis. We borrowed the term Body Copy from the layout and design community, to refer to the main text of a document, as compared to logos, images, title, advertisements, comments etc.

Rather than construct body copy extractors for a potentially infinite set of document types (though clearly this would be needed for a real-world deployment), we created both a PDF and HTML extractor. PDF body copy extraction was made possible leveraging existing Java libraries [4]. To handle the more diverse and complex extraction of HTML body copy, we leveraged the CETR algorithm [49] which extracts body-copy from webpages using HTML tag ratios and multidimensional clustering. Unlike more template driven approaches [7], CETR can extract body-copy from almost any HTML page without knowing layout a priori. While it may produce a few more errors, it will perform on a larger collection of web documents.

Both further stages in our algorithm and end users will need access to this body copy. We therefore automatically extracted sentences and words through use of a parser [48]. Thus we create indexable body copy for future ML analysis, while simultaneously generating HTML with <span> elements around each sentence, allowing those elements to extracted or highlighted later in the interface using JavaScript lookups.

6.2 Stage 2. Identify Discussion Subjects

In order to determine to what degree a sentences’ opinions are Aligned, Divergent from those of the crowd, we must have common subjects on which to compare these opinions. Thus we performed the following process to identify the subjects of discussion within the corpus. In addition, this process calculates a quantitative ranking of the most important discussion subjects to further inform presentation order.

We begin by filtering all body copy for grammatical sentence subject by applying POS tagging3 [3]. With this filtered body copy, we apply Latent Dirichlet Allocation (LDA) topic modeling [34] to extract topics and associated keywords across all documents.

In order explain how we use the outcome of LDA, consider T to be the set of all topics (to . . . t), D to be the set of all documents in our corpus (d0 . . . dl), and W to be the set of all unique words (w0 . . . wn) across all documents. Every document (di) consists of a subset of words from W. Every topic (tj) consists of a subset of words from W, each with a probability P(wj|tj) that if a word is randomly chosen from tj, it is that particular word (wj). In this way, tj is a probability distribution over W:

\[
\sum_{q=0}^{m} P(w_q|t_j) = 1
\]  

and a document (di) consists of a probability distribution over all topics T:

\[
\sum_{j=0}^{k} P(t_j|d_i) = 1
\]

In layman’s terms, every word in a document is contributed by one of the topics.

Thus, LDA gives us the rank (P(wj|tj)) of each word (wj) given a particular topic (tj). However, these word probabilities cannot be compared across topics as they are conditional, and our goal is to rank order all words in W across all topics and documents. Therefore, using Bayes’ Rule, we marginalize out the topics for each word:

\[
P(w_q) = \sum_{j=0}^{k} P(w_q|t_j)P(t_j)
\]

This calculates a rank for each word that is topic independent, comparable across our document corpus. Through this ranking, we now know the most important subjects of discussion in the corpus. To limit the noise from LDA, we only consider the top 500 subjects of discussion.

6.3 Stage 3. Detect Opinion Sentiment

Next we attempt to identify the opinions on each of the important subjects of discussion. We apply a statistical parser[15] to each sentence in each document, generating a full parse tree, a dependence parse tree, and the part of speech for each word. These sentence parse trees are akin to sentence diagramming done in grade school. From these parse trees we can programmatically uncover which modifiers (e.g., adjectives) were applied to any sentence subject.

For each modifier uncovered, we calculate a sentiment score by looking up its value4 in Senti-WordNet[6]. Senti-WordNet is a hand-compiled list covering a large selection of common word modifiers. We rely upon modifier lookup because it does allow us to explicitly associate a sentiment score with a subject. Thus, we can compare the sentiments on a given subject in a given sentence, in a given document to the sentiment on the same subject shared by the crowd. This crowd vs. instance comparison is not possible with holistic sentence-level sentiment calculation.

From our modifier lookup, each word is evaluated on two 0.0-1.0 scales, Positive and Negative5. Thus, for each of the important subjects in Stage 2 (wq), we uncover the set of opinions O(wq) within each sentence of every document where the subject occurs:

\[
O(w_q) = \{o_1(w_q) . . . o_n(w_q)\}
\]

It should be noted that we also identify negations in the parse tree, so we can invert the opinion of the modified subject. For example, in the phrase “the stock is not a good buy,” the subject “stock” is modified by “good.” However there is a negation - “not” - requiring us to swap the opinion’s positive and negative scores. We also wish to note that we dismiss any neural opinion (with a positive and negative score of 0).

2Grouping and clustering articles by subject matters is an active area of research for those in the IR and ML communities, and is outside the scope of this work.
3While an alternative could be Named Entity Recognition[17], entity disambiguation is an open challenge, and we opted to utilize a more conservative approach.
4Modifier lookup does not have word sense disambiguation (context).
5We maintain both positive and negative values because if combined on a single scale, on average, most results would trend towards neutral.
6.4 Stage 4. Calculate Crowd’s Opinions

Next, we aggregate all opinions on each subject over all documents. First, we find all instances of each subject by computing the base form (e.g. removing plurals) of each subject[48] and americanizing all spelling. Thus, for a given corpus, we have a weighted list of subjects from stage 2 and for each subject a distribution of opinions from stage 3.

From these per-subject distributions ($O(w_q)$), we can calculate a mean positive $\mu_{pos}(w_q)$ and negative sentiment $\mu_{neg}(w_q)$, as well as the standard deviation $\sigma_{pos}(w_q)$ and $\sigma_{neg}(w_q)$. We refer to this as the crowd’s opinion on a given subject. For each subject-opinion instance in each document $o_v(w_q)$, we calculate that instance’s opinion distance $F(o_v(w_q))$:

$$F_{pos}(o_v(w_q)) = \frac{|\mu_{pos}(w_q) - \mu_{pos}(w_q)|}{\sigma_{pos}(w_q)}$$

$$F_{neg}(o_v(w_q)) = \frac{|\mu_{neg}(w_q) - \mu_{neg}(w_q)|}{\sigma_{neg}(w_q)}$$

$$F(o_v(w_q)) = F_{pos}(o_v(w_q)), F_{neg}(o_v(w_q)) >$$

By using standard deviation to normalize, we can compare an opinion on any subject to any other subject. All comparisons, therefore, will be a measure of standard deviations away rather than absolute distance (which can vary based on the sentiment distribution variance on each individual subject).

6.5 Stage 5. Score Alignment and Divergence

Next we consider how to calculate an Alignment and Divergence score for each Opinion Sentence and document. We use the term Opinion Sentence to describe a sentence that contains at least one top 500 subject from Stage 2 and at least one non-neutral opinion from Stage 3 on said subject. Not every sentence in every document is an Opinion Sentence. Whereas $O(w_q)$ is the set of all opinions on a given word, we define $S$ as the set of all opinions across all words in all documents:

$$S = \bigcup_{q=0}^{m} (O(w_q))$$

$$\vdash s_r = o_v(w_q)$$

Thus, $s_r$ represents an individual opinion on a specific instantiation of a word. We can substitute little $s_r$ for any $o_v(w_q)$. We can therefore use the following scoring equations to calculate Alignment and Divergence:

$$F_L(s_r) = MAX(F_{pos}(s_r), F_{neg}(s_r))$$

$$A_{opinion}(s_r) = A_{opinion}(o_v(w_q)) \propto P(w_q) \times F_L(s_r)$$

$$V_{opinion}(s_r) = V_{opinion}(o_v(w_q)) \propto \frac{P'(w_q)}{F_L(s_r)}$$

We define $A$ as the Aligned Score, and $V$ as the Divergent score. The subscript opinion connotes the score for a subject-opinion (as compared to a sentence or document).

The scoring equations were based on the notion that both Alignment and Divergence grow as importance of the subject increases. However, while Alignment grows as distance from the crow’s opinion shrinks (hence in the denominator), Divergence should shrink (hence as a multiplier). Further by including the importance of each subject in the equation, both $A_{opinion}$ and $V_{opinion}$ become subject independent, and can be compared or combined with the $A_{opinion}$ and $V_{opinion}$ of another subject (which may have a different weight in the model from Stage 2).

6.5.1 Sentence & Document Level Scoring

We now consider $G$, the set of all subject opinions, $\{s_r\}$, within a given sentence. Thus, we calculate a given sentence’s Alignment and Divergence as follows:

$$G = \{s_r\}$$

$$A_{sentence}(G) = \sum_{z=0}^{n} A_{opinion}(g_z)$$

$$V_{sentence}(G) = \sum_{z=0}^{n} V_{opinion}(g_z)$$

Further, we can define a sentence’s Relevance as a combination of both it’s opinion relationship features:

$$R_{sentence}(G) = A_{sentence}(G) + V_{sentence}(G)$$

Calling out sentences that are highly different or highly similar. Thus Relevance becomes a measure of a broad set of opinions, those that are both Aligned and Divergent.

We now can consider $H$ as the set of all Opinion Sentences in a document. Scoring a document functions similarly to scoring a sentence:

$$H = \{G\}$$

$$A_{doc}(H) = \sum_{y=0}^{|H|} A_{sentence}(h_y)$$

$$V_{doc}(H) = \sum_{y=0}^{|H|} V_{sentence}(h_y)$$

$$R_{doc}(H) = A_{doc}(H) + V_{doc}(H)$$

By using sum, rather than mean, we give more weight to documents that have many opinions, rather than just one strongly opinion.

6.5.2 Classification

We classify a subject-opinion to be Aligned or Divergent as follows:

$$IF(A > V) \Rightarrow Aligned$$

$$ELSE \Rightarrow Divergent$$

regardless if A and V are measuring a subject opinion, sentence, or document.

6.6 Stage 6. Ranking

We utilize two slightly different sorting algorithms for documents and sentences (the two main units of consumption for users of Odin). These two approaches help the Odin UI determine which document or sentence to show to the user in what order.

6.6.1 Opinion Sentence Sorting

Sentences with more superlatives, more subjects, simply more content, generally have the potential to contribute more (or more strongly contribute) to a user’s understanding of the discourse. We therefore order Opinion Sentences on their raw Alignment, Divergence or Relevance scores as calculated by Equations 9 and 10.

6.6.2 Document Sorting

Document length, unlike Opinion Sentence length, have the potential to contain more Opinion Sentences simply because they have more words (not because the document is more opinionated). Consider a corpus where all documents contain at least 1% Divergent
comments. A longer document, which is largely Aligned, may appear to also be very Divergent, simply because of it's length (as compared to a short document which is 95% Divergent, but so short that the sum is smaller than the longer document). We therefore cannot rely upon the raw Alignment, Divergence or Relevance scores as calculated in Equation 11. Moreover, users when they look for documents that are more Divergent or more Aligned, want a document whose opinions are substantially more opinionated (in whichever direction). Therefore we sorted documents by delta scores ($\delta$):

$$\delta_{\text{Aligned}}(H) = A_{doc}(H) - V_{doc}(H)$$

$$\delta_{\text{Divergent}}(H) = V_{doc}(H) - A_{doc}(H)$$  \hspace{1cm} (13)

Thus, documents whose Alignment score is much larger than Divergent score will appear more Aligned, and documents whose Divergent score is much larger than Alignment score will appear more Divergent. The “most” Aligned or Divergent document is the one with the highest respective delta score per Equation 13.

7. INITIAL REACTIONS FROM USERS

An informal study was conducted to understand how well our system and algorithm would support users’ expectations for document corpus exploration with limited time, gauge initial reactions to its mobile interface, and understand how it could be improved. This informal study is not intended as a validation of Odin.

10 users participated (8 male, 2 female), and were shown a demonstration of Odin, its features and capabilities. Given that the majority of opinion based systems are on desktops[23], and focus on summarization rather than document finding and opinion callouts based on similarity with crowd conscious, we were unable to provide a direct comparison to Odin. Though a straw-man, we also showed users a Google News “more sources” page (listing multiple documents on the same news article), and demonstrated how different news/article websites (e.g. CNN, ABC, etc.) rendered articles for mobile devices. Following the demonstrations, participants were given 5-10 minutes to explore Odin. During the informal 20-30 minute sessions, participants were encouraged to think aloud, and comment on any features they liked, disliked, or wish were present. At the end of the sessions, participants were asked to fill-out a brief questionnaire about their experience using Odin.

All were recruited from a large US corporation, had a mean age of 41.1. All of our participants reported using large collections of documents daily or a few times a week. Participants had varying educational backgrounds, both degree level (1 BS, 3 MS/MA, 6 Doctorate) and degree area (5 CS, 2 Statistics, 1 Math, 1 Business, 1 Engineering).

7.1 Feedback & Discussion

Participants were overwhelmingly positive in their feedback regarding Odin with respect to its ability to manage a large corpus:

- "It had a fluid look and feel that allowed for the rapid exposure to a number of articles with little thought about trying to "find" them." - P2
- "As a boss would say... 'give me the five minute rundown,' and you are asking for that, automatically. But now, you don't have to read everything to get that. You don't have to go 'not relevant, not relevant, not relevant.' Well you can, but if you don't have that hour, you can just go through it - which is great!" - P7

and Odin's Executive Summary UI, that quickly brings key opinions within each document to the forefront:

- "Executive summary preview with aligned and divergent key sentences gives me a better overview than the standard abstract or summary." - P10

Very useful that it infers the main ideas and opinions, to give a higher-level understanding of the context of the corpus as a whole before I start to look at documents on a more detailed level." - P8

While the reaction was positive, we explicitly asked users what additional features could be useful to readers on the go. Surprisingly, each participant proposed a different idea:

- A quick tap button for “popular” sources of content (e.g. the author of the document is more reputable or well known)
- A quick tap button for documents that were read by many people viewing this corpus (social popularity)
- See articles that are aligned/divergent from the currently open document (relative alignment)
- Adapt the ranking based on individual user’s viewing preferences or histories across multiple corpuses
- List, on splash screen, the most Relevant, Aligned, and Divergent opinions from the most Relevant, Aligned, and Divergent document (respectively)
- Summary statistics (e.g. number of Aligned opinions on topic A)

As the studies went on, we asked participants what they thought of suggestions we heard from prior subjects. Unexpectedly, participants generally did not feel the others’ suggestions would be valuable. Given that each participant justified their additional feature within their own unique work or situation, we believe that this suggests that the features currently in Odin are universally critical, though additional features may be able useful if turned on based on individual preference.

P1 discussed an interesting perspective of summarization UI (e.g. tag clouds, or summaries) versus Odin’s approach of Alignment and Divergence:

- "This is probably something generalizable, usually you get the official opinion... you see the same thing over and over again, but if someone says something different, now that is interesting so it is important that you find it quickly." - P1

He continued to articulate that traversing a corpus is more than just seeing key words, but seeing relative opinions. In a similar vein, P9 commented on how Odin appears to capture the key features for truly understanding a corpus, beyond that of a simple summary and that the features currently implemented are the vital ones for a smooth interaction:

- "In a world where you have inundation of information, and opinions you could spend 'how many?' hours a day trying to read much less put it together... but this
looks interesting, very interesting, I like it... and I think a lot of these [features] if you don’t have them, you looses a lot. I would say most of [the features] are key features. Cause if you’re not, then your missing too much. What you are looking for is a minimum ante to play the game. If you don’t have this [indicates Odin physically] a lot of these things... its hard enough to take your thoughts and put it in a verbal or written form in the first place, and now the challenge is what did you mean by that. - P9

All users also commented about the elegance and simplicity of the Odin UI. Often these comments were associated with praise that something so simple could despise such a powerful back-end and service especially for a mobile situation when people are on the go:

It appears to allow for personalization of how the information is presented and allows it to meet the demands of the moment instead of dictating what you "should" be interested. - P2

It is important to note that even while the individual algorithmic components in Odin make use of the most straightforward approaches used in the ML and NLP domains, participants explicitly stated that the key sentences were helpful and appeared to be picked out well by the system:

It behaves like a "smart" system. For what I’ve seen, results are always appropriate. - P3

Many users even commented how strikingly different the documents were, and how they immediately saw the distinctions between the three ranking/categories. We strongly believe that this is due to the power of aggregating opinions based on the wisdom of the crowd, combining both subject importance in a topic model with sentiment detection and delta from the mean.

Overall, results from our study provide very promising, but still preliminary evidence that Odin can be a new medium for consuming document corpuses on the go. Encouraged by these results, a formal study is now being conducted.

7.2 Limitations

Given the informal nature of this study, the small N, and the unstructured exploration (rather than explicit task), and qualitative data gathered, the findings should not be interpreted as a validation of Odin itself. However, users reactions do provide a insight into which features they felt were more/less powerful. We can leverage this feedback to better structure a large N study within a real-world context, and improve/aiad Odin’s features.

8. VIDEO OF ODIN IN ACTION

To demonstrate Odin, we have created a demo video viewable online at: http://youtu.be/8_DJteEXk1gI

9. FUTURE WORK

While Odin leverages standard Machine Learning algorithms to perform its analysis, there are many additional topical and sentiments analysis libraries on the “bleeding edge.” As Odin development continues and expands, we aim to continue to integrate more advanced APIs and techniques or sentiment analysis (taking sentence context into account), keyword extraction (with n-gram phrases and Named Entity Recognition) and topic modeling (combining supervised and unsupervised techniques). As stated above, however, participants’ initial reaction to Odin’s performance was quite good.

Another direction of future work would be the inclusion of social aspects of corpus reading as suggested by many of our participants. While Odin focuses on the written corpus itself to weight rankings, we wish to explore how we can integrate user’s view habits into the algorithm to “bubble up” popular articles and opinions. Further, we wish to additionally explore taking source “credibility” into the algorithm’s weight. Whether the source is the enterprise hierarchy of the author, or the veracity of the journal/author from more public domains, we believe source can play a powerful role.

Lastly, we are currently undertaking a long-term ecological valid study of Odin, integrating it within a corporate or legal environment. This would allow us to better understand how Odin could quantitatively impact job performance.

10. CONCLUSION

The digital age and the resulting increased informational awareness while human time and attention is becoming scarce. This problem is exacerbated by the growing ubiquity of mobile devices as a major information consumption platform, in that they have smaller screen real-estate and mobile device users have far less time to consume content. In this paper we presented Odin, a novel system and algorithm that leverages the “wisdom of the crowd” to create an interface into a users’ document corpus from three distinct perspectives; Relevance, Alignment and Divergence. Thus, rather than simply seeing a summary of a corpus keywords, Odin lets a user quickly access key opinions and documents by their relation to the crowd’s conscious. In order to facilitate this interaction Odin grounds many UI decision in a breath of literature from many disciplines. A series of initial user interviews suggest strong support for the quality and utility of a system like Odin within real-world scenarios. Thus, the foremost contribution of this work is an interaction model and algorithmic approach that allows users to place each document and opinion within the larger corpus of opinions while “on the go.”

11. ACKNOWLEDGEMENTS

12. REFERENCES
