

Intention Insider: Discovering People's Intentions in the Social Channel

Malu Castellanos
HP Labs
Palo Alto, CA
malu.castellanos@hp.com

Meichun Hsu
HP Labs
Palo Alto, CA
meichun.hsu@hp.com

Umeshwar Dayal
HP Labs
Palo Alto, CA
umeshwar.dayal@hp.com

Riddhiman Ghosh
HP Labs
Palo Alto, CA
riddhiman.ghosh@hp.com

Mohamed Dekhil
HP IPG-IWS
Palo Alto, CA
mohamed.dekhil@hp.com

Carlos Ceja
HP Labs
Palo Alto, CA
carlos.ceja@hp.com

Marcial Puchi
HP Labs
Palo Alto, CA
marcial.puchi@hp.com

Perla Ruiz
HP Labs
Palo Alto, CA
perla.ruiz@hp.com

ABSTRACT

The rapid proliferation of online forums has made it possible for people to share their intentions, wishes and experiences by posting comments with the aim of getting advice from other members of the forum. Extracting intentions from these comments provides valuable insight for companies who can exploit it to get a competitive edge. However, given the very large amount of this kind of online comments, manually extracting intentions is impractical, time consuming and expensive. Companies need tools that analyze the text to extract intentions and details about them. In this paper we propose to demo one such tool called *Intention Insider* which has been developed at HP Labs in close collaboration with business units and a few selected customers. The tool can ingest content from online forums or from uploaded files and quickly sift through very large amounts of comments to extract intention information. This information is loaded into a data warehouse to be correlated with other structured data and queried to produce interactive reports and dynamic visualizations that facilitate its exploration at detailed and aggregate levels.

Categories and Subject Descriptors

H.4.m [Information Systems Applications]: Miscellaneous. H.3.5 [Information Storage and Retrieval]: Online Information Services – *Web-based services*.

General Terms

Algorithms, Experimentation.

Keywords

Intention analysis, text mining, information extraction, social media, visualization.

1. INTRODUCTION

With the increase in social networking websites, forums, blogs, and similar Internet websites people are more and more willing to

share information regarding their intentions, wishes, opinions, and experiences. In particular, forums present an opportunity to share posts related to a topic to get input about them from other members of the forum. For example, forum participants may share their intentions regarding a vacation they plan on taking to solicit advice from others.

Companies have been quick in realizing the potential value of the information expressed in the social media channel which once analyzed, may be utilized to capitalize on an understanding of consumers' thoughts, perspectives, needs, intentions and wishes. Specifically, such an understanding can be used to provide better products and services to consumers, enhance customer relations and create personalized target marketing. Thus, the ability to analyze this information in a timely manner enables companies to gain a competitive edge in a fast dynamic market.

In spite of the fact that there is an ever and rapidly increasing amount of information shared in the social channel, companies cannot capitalize on its value unless the information can be extracted in some way that is useful. In the case of online forums where people express their plans and intentions --which is the focus in this paper-- companies typically have a team of people manually sifting through this content to extract intentional information that can subsequently be exploited. Such is the case of businesses in the hospitality and entertainment industries that exploit the contents of vacation related forums, or the case of hardware companies that manually comb on-line surveys about people's intentions to buy or upgrade equipment.

With the incalculable amount of online text available within the innumerable online forums, a company would find it impossible to manually process the text in search for statements of intention. It would be too expensive for a business entity to create a team of people for analyzing this online text. Further, the pace at which this intention extraction could be performed manually could not cope with the pace at which the streams of messages are created, and often the online text would become "stale" in that the statements of intention would quickly become outdated or simply the opportunity to capitalize on their value would have passed. It is therefore essential to equip companies with tools based on state of the art technology to ingest online content from sources of interest, analyze this content and extract valuable insight, in this case people's intentions, that can be presented to business people in an intuitive way that facilitates its exploration.

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.

EDBT'12, March 26–30, 2012, Berlin, Germany.
Copyright 2012 ACM 978-1-4503-0790-1/12/03 ...10 00

In this paper we propose a demo of our intention analysis tool, *Intention Insider*, which is part of the Live Customer Intelligence project being developed at HP Labs. At the core of the tool natural language processing (NLP) and text mining techniques are used to extract intentions and their attributes. This information is stored in a database to be queried upon requests from the user via a dashboard that presents the information on interactive visualizations. Throughout the development of this tool we have had close interaction with HP business divisions and selected HP customers whose requirements have shaped the functionality of the tool.

In the rest of the paper we give an overview of the architecture and techniques (Section 2) and describe the demo (Section 3).

2. OVERVIEW AND ARCHITECTURE

In this section we provide a brief overview of the *Intention Insider* system architecture shown in Figure 1. *Intention Insider* is intended to be a cloud service that provides REST web service interfaces. It is architected to operate on online content such as online forums as well as stored content such as call center surveys. Content ingestion adapters pull data from instances of online forums that is fed to the Intention Processor where it is preprocessed and cleaned before natural language processing and information extraction techniques are applied to it. Once intention information is extracted, it is loaded into a data warehouse to be correlated with other structured data and queried in response to users requests from the frontend. The results are displayed as interactive reports and dynamic charts for an intuitive visualization at the aggregated and individual level.

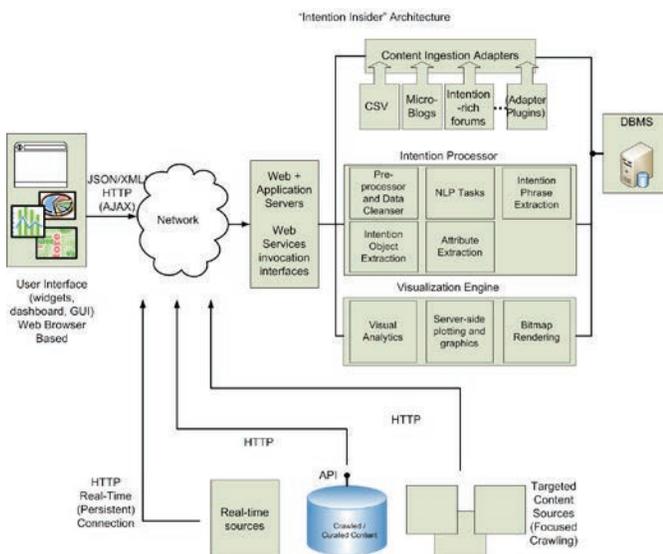


Figure 1. Intention Insider System Architecture

Intention analysis as its name suggests is the task of analyzing the free form text in documents where people express their wishes, intentions and plans to extract structured data about the intentions expressed in natural language. To the best of our knowledge there are no tools for doing this kind of analysis nor the literature has reported work to extract information about intentions. It might be perceived as similar to sentiment analysis [1],[2] since it also extracts insight into what people expresses but apart from some common building blocks like preprocessing and NLP, the core

techniques are well different. In sentiment analysis the goal is to assign positive, negative or neutral polarity to opinionated text to understand, for example, what customers liked or disliked about their stay at a given resort. In contrast, intention analysis tries to understand what are the customers plans with regards to the resort. In the rest of this section we give an overview of the underlying techniques that systematically analyze textual corpora to extract people’s intentions and some details.

The method consists of 2 main steps:

- a) identifying the intention phrases
- b) extracting the intention components.

An *intention phrase* are those phrases that clearly mark that there is an intention, phrases like “would like to see”, “are planning a trip”, “thinking about doing”, indicate that the comment is about an intention to see something, to make a trip and to do something, respectively. An intention phrase can have different components but it is commonly formed by an *intention verb* and another particle like a preposition or an article, like in “like to”, “planning a”, “thinking about”. The intention verb may be associated to an *intention action* which as its name suggests is the action intended, like “see” (in “like to see”) or directly to an *intention object* which is the object of the intention like “trip” (in “planning a trip”). The intention action is typically associated to an intention object like “Mickey” (in “would like to see the Mickey show”). The *attributes* of the intention include the intention object as well as complementary information that give details about the intention.

For example, given the following intention comment:

“Hi, we are planning a trip in June with our 7 year old daughter and 13 year old son. Can you tell me what princesses are at Cinderellas castle for the fairy Tale Lunch? We have done dinner there on our last trip and would like to see more princesses.”

One intention verb is “are planning” for which the intention object is “a trip”, the other intention verb is “would like to” for which the intention action is “see” whose intention object is “princesses”. The attributes of the intention comment are:

- Ages: 7 and 13
- Date: June
- Party size: between 3 and 4 (since the people mentioned are children, at least an adult should be in the trip too)
- Attraction: Cinderellas Castle

The extraction of the intention phrase and its components, except the attributes, uses an NLP-based technique, whereas the extraction of the attributes uses information extraction techniques (IE). Next we explain how our method works to extract these elements about the intention.

2.1 Intention Phrases

Typical intention phrases are of the form:

[intention verb] + [action verb] + [intention object]

e.g. We plan to stay at Pop Century.

Intention Verbs and Action Verbs (either one or the other) are sometimes optional in natural language when expressing intention. To identify intention phrases and their associated action verbs, we used a bootstrapping approach where a list of seed intention phrase patterns was created, and matched against the corpus to find text that conforms to these patterns. The highly frequent intention verbs and action verbs discovered from this process were then used to find other intention phrases. We

repeated the process until no more new intention phrases were found. Bootstrapping increases recall but affects precision since it returns some phrases that are not intention oriented. However, this occurs mostly on the long tail of the phrases distribution, which does not affect the intention analysis where the focus is on the statistically relevant intention phrases. In addition, the visualization, aimed at exploring results, facilitates the identification (and elimination) of those cases in which incorrect phrases are statistically relevant.

The seed intention phrases patterns we used are the following:

- Verb + Infinitive: <verb> <TO> <verb>
(e.g. *We want to visit Teppan Edo at Epcot*)
- Verb Pairs + Preposition: <verb> <PREPOSITION> <verb>
(e.g. *We are thinking of staying at the Yosemite Cabins*)
- Only Action Verb: <action_verb>
(e.g. *We are going to see the Magic show.*)
- Only Intention Verb: <intention_verb>
(e.g. *I am planning a deluxe solo cruise*)

To match the above patterns with the corpus first the text in the online forum comments needs to be preprocessed by passing each comment through the following NLP pipeline:

- Sentence detection- to split the text into sentences. A sentence is the scope of an intention phrase.
- Tokenization – sentences a break down into tokens
- Parts-of-speech (PoS) tagging – each token is labeled with its PoS such as noun, verb, adverb, etc. This step is required because our patterns use PoS tags.

Then the labeled corpus is matched against the specified intention phrase patterns to find new patterns. A small sample of such detected patterns present in the corpus of an online forum along with their frequencies is shown below:

331 going to be	203 planning on going
218 like to know	189 thinking of staying
159 want to make	156 planning on doing
125 have to pay	134 thinking about doing
82 planning to go	92 planning on eating
76 want to see	73 thinking of taking
73 trying to decide	56 thinking about staying
66 need to bring	46 go about getting

Based on the intention phrases discovered in the text, common (high frequency) intention verbs (e.g. like, want, planning, need) and action verbs (go, stay, taking, see) are identified. These intention and action indicators can be used for the “only action verb” and “only intention verb” patterns.

In addition, the level of intention (in terms of certainty) can be classified based on the intention verb; this allows for a deeper understanding of the intention expressed. (E.g. “thinking of going” expresses weaker intention certainty than “want to stay”).

The next step is to analyze sentence structure to determine the *object* of the intention. Different techniques such as grammar based chunking, and dependency parsing can be used to determine the object but they are expensive. Fortunately, most often the object comes right after the intention phrase and our experiments revealed that using the noun that comes after the intention phrase gives very high recall and precision (above 0.9) on the object extraction.

Extracting the intention phrase and the intention object and the timeline allows us to understand the intention expressed, and also on an aggregate level, obtain business intelligence on what “intention objects” are the most frequent, which co-occurrences of action verb and intention object are statistically significant, and the temporal characteristics of these intention objects.

2.2 Intention Attributes

Once the intention phrase has been recognized, the next step is to extract all the relevant components of the intention. This is done by creating information extractors for the different types of entities to be extracted. For the example given earlier, there is an extractor for **ages**, another for **dates**, another for **party size** and finally another for **attraction**.

Some extractors are created manually others automatically. Our manual extractors correspond to attributes that are an enumeration of specific values or that can be recognized with a small number of expressions. Henceforth, two following two methods are used:

A. Constructing a lookup table with names to be recognized. For example, for the **attraction** extractor, a list of the attractions of interest could be given.

B. Constructing a rule set where in each rule the conditional part is a regular expression formed by tokens (or generalizations of them) in the surrounding context of the entities to be extracted. For example, for the ages extractor, one possible regular expression could be “age{s} [num] {,}and [num]}” which means that the pattern to be matched is the word “age” or “ages” followed by a number and optionally a repeating group of comma or “and” followed by another number. This expression constitutes the conditional part of the rule formed by the pattern to be matched. The action part of the rule is the extraction of the numeric parts of the pattern which are the ages.

To build extractors automatically we use two techniques:

C. Rule induction where the patterns of the rules are learned using a genetic algorithm that is trained to recognize the attribute context.

D. Conditional Random Field where a statistical model that is more robust to noise and variations is trained on a labeled set.

Both techniques require a training set which consists of a representative set of documents manually labeled with the entities to be extracted. Given the variability of the contexts in which a given entity appears it is common that an entity has more than one associated extractor. Once this information has been extracted and stored in a data warehouse, it can be queried according to the user requests coming from the frontend to give detailed insight into people intentions.

3. DEMO DESCRIPTION

The purpose of the demo is to show the main features of our intention analysis tool *Intention Insider*. The tool can be used in standalone mode or as an additional backend service in our Live Customer Intelligence platform [1]. The user interacts through a frontend that has four tabs and we plan to run the demo by letting the user (audience) go through each of them to experience the easiness, flexibility and power of the tool. Next we describe how we plan to run the demo:

i) Settings tab – through this tab the user enters the configuration information of the data source which can be either an online

forum or a file. He can also activate the intention analysis from this tab so that the tool ingests the data and proceeds with the analysis as explained in Section 2. This tab also lets the user to upload a logo to be displayed on the different tabs.

ii) Interactive report - once the analysis finishes and the results are loaded into the data warehouse, the user is notified and he can choose to go to any of the other three tabs. First he will go to the one that shows an interactive report like the one in Figure 2. The upper part of the report has a tag cloud that shows the intention objects as tags. The size of a tag is proportional to the frequency of the corresponding object over all the comments of the dataset (i.e., a measure of the importance or interest that people has in that object). The user can click on a tag in the cloud to retrieve the comments whose intentions have the corresponding object as target (middle part of Figure 2 where the object mention is colored on each comment) and to obtain information about the attributes (i.e., details) of the intentions involving that object (lower part of Figure 2). The attribute information is presented as bar or pie charts displaying the distribution of the values of each attribute. In the example in Figure 2 the pie charts display the ages, party size and month of the intention to visit Walt Disney World (whose tag in the tag cloud has been clicked), where all this information has been automatically extracted from the comments.

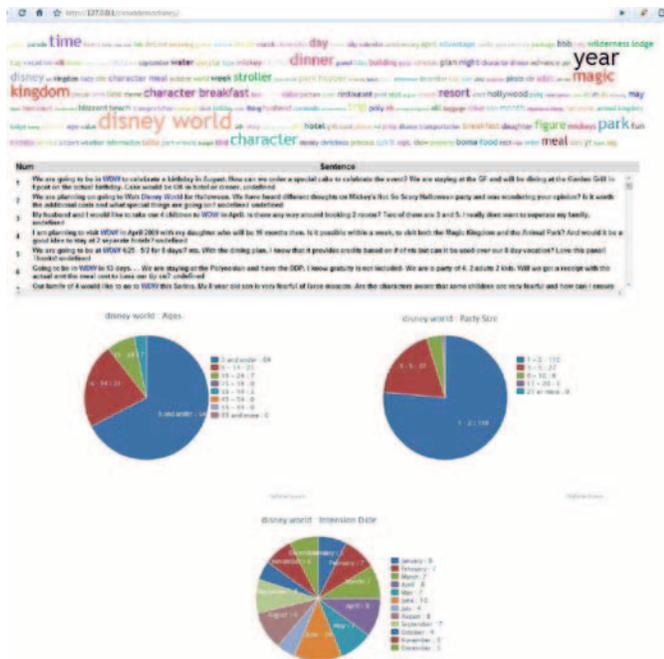


Figure 2. Interactive report

iii) Form-based querying - the user can also query the database through an interactive form whose fields correspond to the parameters of useful predefined queries. The interactive form is displayed on the Querying tab.

iv) Explorer – this tab lets the user navigate through the different components of the intentions to get a good understanding of the significance of co-occurrences of intention verbs with actions verbs, action verbs with objects, and in general any co-occurrence of different components of the intention phrases in the corpus. As shown in Figure 3 on the left side of the panel there is a menu of filters of different categories corresponding to pattern types,

patterns and objects (collectively called elements in this section) found in the corpus that allow the user to explore those of interest. When the user makes a selection the filter boxes get updated accordingly and a visualization of the selected element and its children is presented in the form of a graph where each node corresponds to an element explored so far and its children. For example, if the user wants to see all the actions of “planning to” type of intentions, he selects the filter or the node [intention verb] TO [action verb] and when the children of this filter appear he selects the planning to node or filter and all the action verbs of intention phrases that have ‘planning to’ are displayed as nodes with links to the planning to node. At the same time all the comments matching the filter selections and the children of the lowest level selected filter are shown on the bottom part of the panel with color coded highlighted words. In the graph the size of a node is proportional to the frequency of its corresponding element in the corpus and its color corresponds to the element category. The user can fold or unfold each node by clicking on it or on the filters and the layout dynamically readjusts while at the same time the filter boxes get updated with the elements. There are also different sliders that let the user configure the visualization layout to accommodate his particular needs.

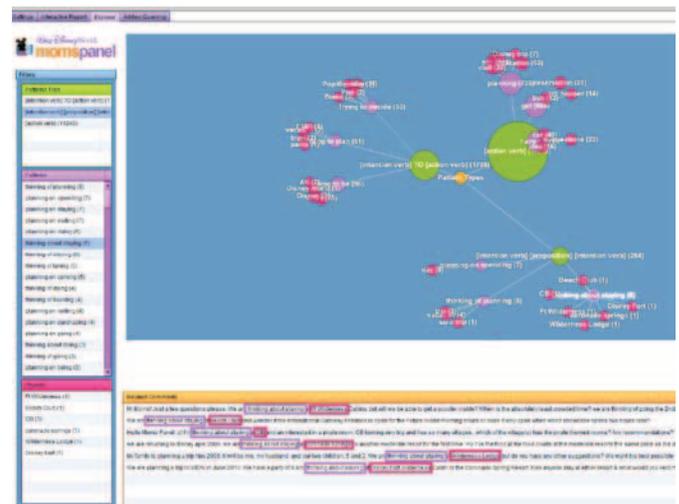


Figure 3. Explorer with filters and visualization

4. CONCLUSIONS

Intention Insider is an intention analysis tool that has been successfully used to understand people’s (and in particular, customers and consumers) intentions published in social media. The demo will show its main features as applied to online forums, and its ability to effectively extract the intention components and present them in a way that facilitates the exploration to enable actionable insight.

5. ACKNOWLEDGMENTS

We want to thank the reviewers for their insightful feedback.

6. REFERENCES

- [1] M. Castellanos et al. LCI: A Social Channel Analysis Platform for Live Customer Intelligence. SIGMOD 2011.
- [2] B. Pang, L. Lee: Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval, 2(1-2).