

1. **TITLE:** Personality based sentiment analysis of textual information written in natural language

2. **IDENTIFICATION OF INVENTOR(S)**

Identify the key inventor to whom questions can be directed. For each inventor, provide the following:

A.	Full legal name	Dheerender Velu Palani
	Employer	IBM
	Work email	dvelu@au1.ibm.com
B.	Full legal name	Josh Andres
	Employer	IBM
	Work email	josh.andres@au1.ibm.com
C.	Full legal name	Lida Ghahremanlou
	Employer	IBM
	Work email	lida.ghahremanlou@au1.ibm.com
D.	Full legal name	Nebula Alam
	Employer	IBM
	Work email	anebula@au1.ibm.com
E.	Full legal name	Rajib Chakravorty
	Employer	IBM
	Work email	rachakra@au1.ibm.com

3. PROBLEM ADDRESSED BY THE INVENTION

Describe the problem or need addressed by your invention. What problem does your invention solve, or what need does your invention fulfill?

In today's world there is an explosion of information and a lot of this is produced as a result of human entered data in natural language. On the internet, the information generated may be as result of social media activity or in the form of user reviews on products / businesses or as blogs. In an office set-up, large amounts of similar information is created to communicate between office workers, primarily using email and chat systems.

Many a times the emotions in the information produced this way can be ambiguous and not always comprehensible. The personality of person and the choice of words used to construct sentences and paragraphs can have a big effect on how the information is captured and how it can be understood by the reader. The reader could misunderstand or misinterpret what is being communicated or sometimes even find it hard to fully comprehend.

Our invention aims to address this problem by performing sentiment analysis on the written text and co-relating that with the personality of the information creator (writer) and thereby resulting in a Personality Based Sentiment (PBS) Analysis of data.

Whilst the algorithm we have invented can be used in any scenario where the information is in an unstructured natural language format and is produced by a human, we have chosen to demonstrate the application and implementation of our algorithm on emails. The rest of the document describes the details using a typical office email system.

3.1. PROBLEMS ADDRESSED IN EMAIL SYSTEMS

Email systems are the most widely used communication mechanism today in both, professional and personal situations. They have been around for over two decades but not much has changed in the way they work and the information they provide to the user. Our invention can be used to address two problem areas associated with email current systems:

- **Ambiguity or Vagueness** in the information carried by the email: *There are times when the information that is sent in the email is ambiguous due to the choice of words used by the sender and may also be influenced by the sender's personality. Our invention aims to alleviate this problem by extracting personality based sentiments from the email and its sender and presenting that to the email receiver.*
- **Prioritizing emails based on sentiment level:** *In current email systems, the receiver is not able prioritise the emails based on the sentiment or emotion level contained in the email, without having to read the entire email body (and even*

then the perception varies due to the difference of personality between the sender and the receiver). Our invention aims to alleviate this problem by presenting the personality based sentiment (PBS) score against each email as preview in the inbox. The user can prioritise for example, 'unhappy' emails from customers over 'happy' emails.

4. STATE OF THE ART

Please list and briefly describe the products, publications, patents, and other works that are most closely related to your invention. If any of these works solve or address the same problem, what are the drawbacks of the known solution(s)?

Nigam and Hurst <http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PG01&p=1&u=/netahtml/PTO/srchnum.html&r=1&f=G&l=50&s1=20060069589.PGNR>. Present a system for performing topical sentiment analysis on electronically stored communications, using NLP shallow techniques. This approach polls topics and provides sentiment distribution. It does not take the user personality into account.

Ham (<http://brevets-patents.ic.gc.ca/opic-cipo/cpd/eng/patent/2767691/summary.html>) A computer implemented method of collecting, exchanging and displaying both real-time and historical sentiments on a topic. This system concentrates on exploring topics, rather than sentiment in relation to the sender's personality, which is our area of focus.

Rehling and Dignan <http://patft.uspto.gov/netacgi/nph-Parser?Sect2=PTO1&Sect2=HITOFF&p=1&u=/netahtml/PTO/search-bool.html&r=1&f=G&l=50&d=PALL&RefSrch=yes&Query=PN/8463595> Presents a system to analyse multiple parts of a body of text that overlap, for example a positive statement, followed by a negative statement (traditionally in NLP positive and negative would cancel each other), however with this system the creator aims at improving the relation of this statements within a body of text. Personality is not taken into account.

Cheng, (<http://bit.ly/1UiZRX4>) presents a system aiming at building user models of personal interest by taking into account user engagement data + using the sensors of mobile device to read sentiment and correlate with the data explored. Our system differs, one area is that our system analyses textual data to determine the sentiment.

Patrick W. Fink, Kristin E. McNeil, Philip E. Parker, David B. Werts (<http://bit.ly/1KBhwD6>) present a dashboard system in which the sentiment from user one is shown next to their inputs, and progressively taking into account how the conversation evolves to assign a sentiment to the next input. This system does not take into account the personality of the user.

A patent held by IBM, Dhruv A. Bhatt (<http://bit.ly/1O4Ui7n>) present a system that aims to

determine the sentiment of an electronic communication. This system relays in identifying sub construct of the email, adding the different sub construct scores and ranking a sentiment to be presented. The system does provide email sentiment ranking, however it does not consider the personality of the user in relation to the email content, this correlation provides the opportunity to offer a more accurate sentiment. This system does not touch on assigning sound, filtering by sentiment, or playing a sentiment sound forecast.

5. DESCRIPTION OF THE INVENTION

5.1. SUMMARY

The email analyzer uses the sender's email id to identify the person and then carries out personality analysis based on their social media activity. Social media can be public platforms like Twitter and/or Facebook or can be internal enterprise social media or collaborating platforms.

Next, the email analyser extracts the textual information from the email body and runs a sentiment analysis on that. The results from personality analysis and sentiment analysis is inputted into a custom algorithm that will correlate the two results and generate a final Personality based Sentiment score (PBS Score). The final score is rendered using simple emoticons (smileys) based on the overall sentiment or tone.

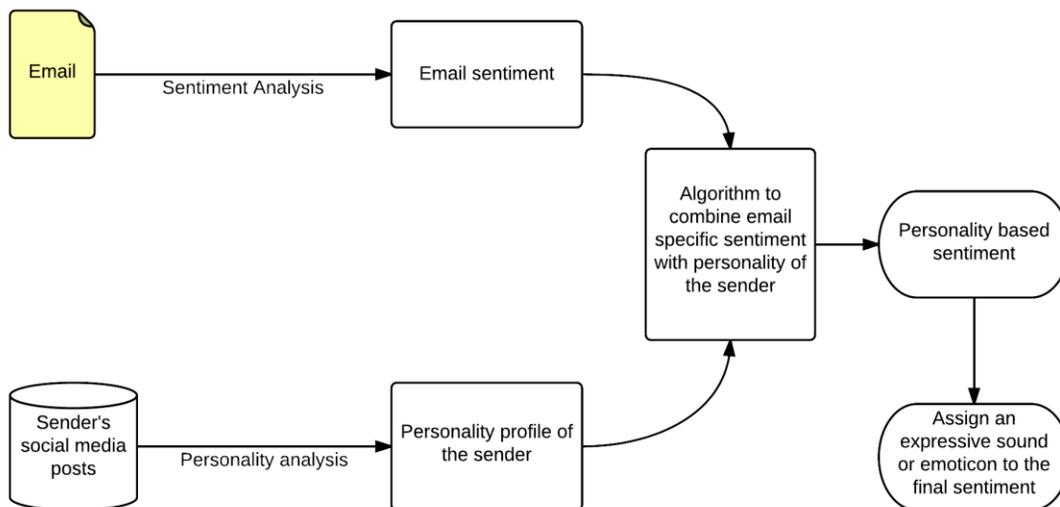
The emoticon is pinned to the email and presented in the inbox view. Each emotion is also complemented with a human like sound that represents the emotion and can be played to "listen to the emotion". Example, a laughing sound represents a very happy and positive PBSA score.

The implementation makes use of cognitive services provided by IBM Watson Services and this document does not provide any information on how these services work. Please refer to IBM Documentation available on [here](#).

5.2. DETAILED DESCRIPTION

Provide a detailed description of what your invention is and how it works. What are the components (or steps if the invention is a method) of the invention and what is the function of each component? How do the components interact with each other? What are the novel aspects of your invention? Include sketches, drawings, photographs, etc. to help illustrate your invention. (Drawings may be attached on a separate sheet or included in the detailed description)

5.2.1. COMPONENT INTERACTION MODEL



The sentiment of the email is first extracted from the email text. But we cannot claim that to be the exact emotion of the sender, as sentiment is only a short-term representation of their emotion and is affected greatly by their long term emotional state, which is their personality. In order to extract the personality of the sender, his/her social media contents are analyzed to obtain insight about their personality. Then the personality and the immediate sentiment is combined algorithmically to obtain a personality-biased sentiment of the email, which can then be represented using a sound or an emoticon.

5.2.2. IMPLEMENTATION & ALGORITHM

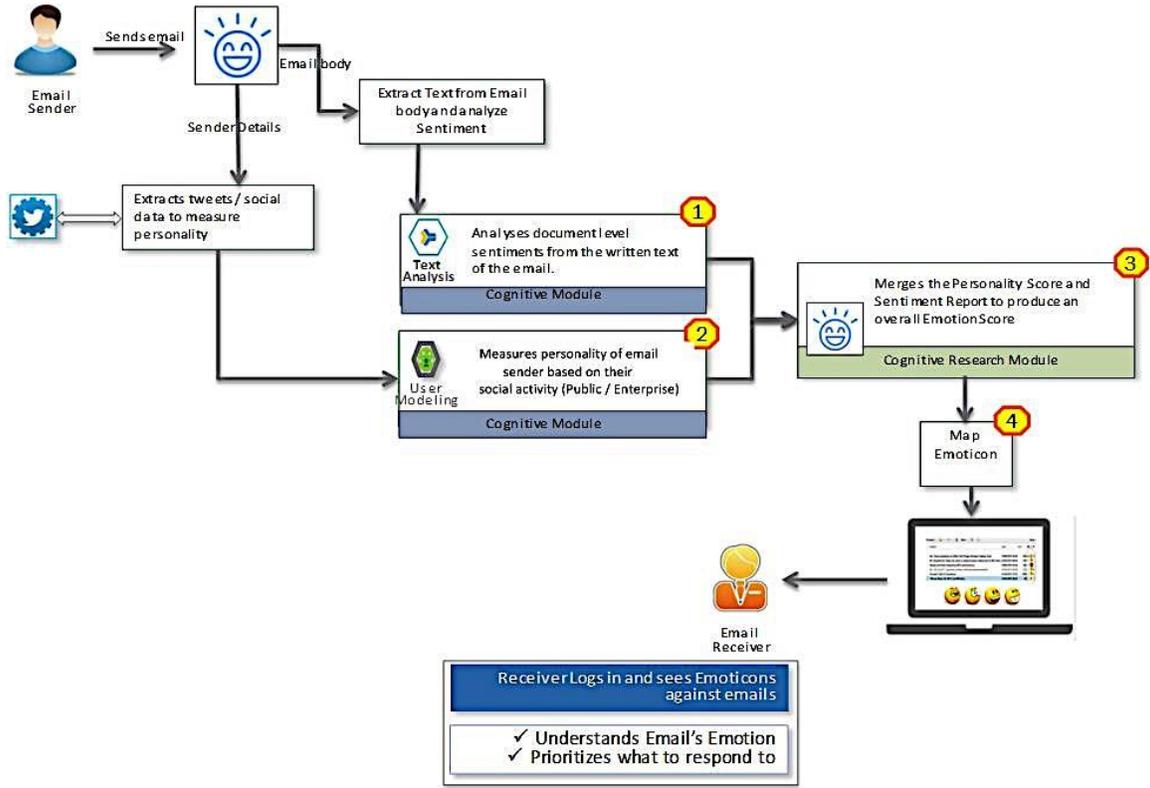


Figure 2: Process Steps

As shown in Figure 2, the application has 4 keys steps:

1. Sentiment Analysis of the Email Body:

The first step in the process is to extract the email body (text) and carry out a sentiment analysis on that information. We have used the IBM Alchemy Sentiment Analysis Service which uses Natural Language Processing based technique and provides easy-to-use mechanism to identify positive/negative sentiment using within any textual information.

More information on how the service API works and the response formats can be found on IBM Alchemy's website at the link - <http://www.alchemyapi.com/api/sentiment-analysis>

The output of this steps is a Positive or Negative sentiment score.

2. Producing Personality Profile of the Sender:

The next step is to gather sender information based on the sender's email address, name and contact and then determine their social media data. Based on social media activity we can generate a personality profile. We have used IBM Watson™ Personality Insight Service to carry out this step.

The IBM Watson™ Personality Insights service provides an API that enables applications to derive insights from social media, enterprise data, or other digital communications. The service uses linguistic analytics to infer personality and social characteristics, including Big Five, Needs, and Values, from text. The output of Personality Insights is the three types of personality models: Big5 (Big Five), Basic Needs and Values, which each has various facets. More information on the service and the science behind it can be found at this link – <http://www.ibm.com/smarterplanet/us/en/ibmwatson/developercloud/doc/personality-insights/>

Personality modelling originated from psychology. Multiple studies ([Green & Fisher, 2010](#); [Rotter, 1966](#); [Ziemkiewicz et al., 2013](#); [Mairesse et al, 2013](#)) have focused on the influence of personality factors on an individuals' behaviour and decisions. In particular, several researchers found that variations in word usage in writings such as blogs, essays, and tweets can predict aspects of personality ([Fast & Funder, 2008](#); [Gill et al., 2009](#); [Golbeck et al., 2011](#); [Hirsh & Peterson, 2009](#); and [Yarkoni, 2010](#)).

After studying the literature, among all the traits that are extracted and analysed by this service, we chose 19 facets that are related to communicate with other people for the purpose of this innovation as shown below green facets have the positive score and red facets have the negative score.

Personality Trait	Facet (Variable)
Big5: Agreeableness	Altruism, Cooperation, Trust
Big5: Conscientiousness	Dutifulness
Big5: Extraversion	Cheerful, Friendliness, Gregariousness
Big5: Neuroticism	Anger, Anxiety, Self-consciousness, Immoderation, Vulnerability
Need	Harmony, Ideal, Structure, Love, Practicality
Values	Self-transcendence, Conservation

Table 1: Personality Traits

We then used these facets to generate the algorithm of the sentiment analysis with personality profile.

3. Personality Based Sentiment Analysis Algorithm :

We defined an algorithm to combine the sentiment analysis with personality profile to obtain the Personality Based Sentiment score (PBSS) of the email

In general, the relationship can be depicted as

$$PBSS = f(\textit{Sentiment}, \textit{Personality})$$

The function $f(\cdot)$ can have any form (linear, non-linear, machine learning and/or statistical modeling based).

In our implementation, the final score was obtained as an weighted average of the positive attitude of the personality profile of the and the sentiment analysis of the email content. The relationship and six types of categories for the final score is given below:

Final Emotional Score = 0.3 * Positive Attitude + 0.7 *Sentiment Analysis

```
If (Final Score <0.2)
Category 6 (Very tense)
Else if (0.2< Final Score <0.3)
Category 5 (More tense)
Else if (0.3< Final Score <0.5)
Category 4 (Tense)
Else if (0.5< Final Score <0.6)
Category 3 (Neutral)
Else if (0.6< Final Score <0.8)
Category 2 (Happy)
Else if (0.8< Final Score)
Category 1 (Very Happy)
```

4. Map PBSS to Emoticons & Compute Sentiment tone:

The final step in the process is to map the PBS Score (PBSS) to emoticons that represent the level of sentiment and then compute an overall tone. We used the 6 emoticons to represent the final 6 PBS scores from previous step

Final PBS Score	Emoticon
-----------------	----------

Very tense	
More Tense	
Tense	
Neutral	
Happy	
Very Happy	

The sentiment forecast of the day is obtained by performing various voice representations of the overall sentiment of each email. The system provides a button for the user to listen to the overall sentiment or the “tone of the day”.

RESULTS AND ADVANTAGES

What are the results and advantages of using your invention? How does it work differently from other devices or processes that accomplish the same purpose?

During testing of our web email inbox implementation of the idea, we have noted significant differences between personality-biased sentiment and raw email sentiments. After taking the personality of the sender into account, the sentiment output of our system represented the magnitude of the emotion expressed in the email better.

In summary, the benefits of our inventions compared to existing solutions are:

- *Sentiment ranking as part of the email preview, in the form of an emoticon.*
- *Filtering / Sorting of emails based on Sentiment (PBSA)*
- *Play emotion sound*

[Please note that differences with some other systems are described in section 4.0](#)