The Personality of Venues: Places and the Five-Factors ('Big Five') Model of Personality

Vlad Tanasescu, Christopher B. Jones, Gualtiero Colombo, Martin J. Chorley, Stuart M. Allen, Roger M. Whitaker
School of Computer Science & Informatics,
Cardiff University,
UK, CF24 3AA
{v.tanasescu, c.b.jones, g.colombo, m.j.chorley, stuart.m.allen, r.m.whitaker}@cs.cardifff.ac.uk

ABSTRACT
Venues are often described by their type and characteristics, while their level of appreciation by users is indicated through a rough score (star rating). However the judgement on a particular venue by an individual may be strongly influenced by the individual's experience and personality. In psychology, the five-factor model of personality, or 'Big Five' model, describes an individual's personality in terms of *openness*, *conscientiousness*, *extraversion*, *agreeableness* and *neuroticism*. This work explores the notion of 'personality of a venue' by reference to personality traits research in psychology. To determine the personality of a venue, keywords are extracted from reviews of venues, and matched to terms indicative of personality traits dimensions. The work is completed with a human experiment where participants qualify venues according to a set of personality descriptors. Correlations are found between the human annotators and the automated extraction approach.

Categories and Subject Descriptors
H.1.2 [**Information Systems**]: Models and Principles – *User/Machine Systems.*

General Terms
Algorithms, Measurement, Experimentation, Human Factors.

Keywords
Five Factor Model, place, personality, reviews, recommendation.

1. INTRODUCTION
Location based Web services have become popular among Internet users, who can leave reviews, or 'tips', for venues they visit. Unlike edited expert reviews, which may be structured according to comprehensive, consistent and objective checklists of features, user contributed reviews are generally short texts reflecting the 'perception' or 'feeling' of a place, usually accompanied by a rating (stars). The aim of this research is to capture this underlying 'feeling' of a place, or personality of a venue, in a way related to personality research in human psychology. This representation of a place that depicts it as a projection of (different) users perceptions is in line with the notion of 'sense of place', which emphasises the characteristics that make a geographical location special or unique from a human point of view [1].

Personality research in psychology is based on the so-called *lexical hypothesis*, stating that the personality characteristics that are most important in peoples' lives will eventually become a part of their language, and that more important personality characteristics are more likely to be encoded into language as a single word. The lexical hypothesis was invoked to derive five broad domains or dimensions of personality, where personality descriptors of individuals cluster together after factor analysis. Dimensions of this Five Factor Model (FFM), also called 'Big 5' personality traits, are usually represented by roman numerals (I-V). For mnemonic purposes the acronym OCEAN is also used. OCEAN stands for descriptions of the five dimensions as *Openness* (inventive or curious vs. consistent or cautious), *Conscientiousness* (efficient or organized vs. easy-going or careless), *Extraversion* (outgoing or energetic vs. solitary or reserved), *Agreeableness* (friendly or compassionate vs. cold or unkind) and *Neuroticism* (sensitive or nervous vs. secure or confident). According to the FFM, everyone's personality can be described with some level of confidence along these dimensions.

The research presented here assumes that venues can also be described using human personality dimensions, due to the personality of the individuals that frequent the venue, or because of characteristics of the venues. This usage is also embedded in language, which sometimes uses the same vocabulary to describe place and human personality. For example an "arty" place is expected to have a higher Openness (O) value than a "traditional" pub. A nightclub would be expected to score high on Extraversion (E), while a "quiet" coffee place should score lower on that personality dimension. An accurate qualification of venues according to their personality should provide insight for better place recommendation, as well as constitute a step towards the computational capture of the 'sense of place'.

The approach presented here is based on reviews collected from location based Web services. Reviews are parsed and matched to adjectives characteristics of personality traits dimensions. A human experiment is presented which is used as proof of concept as well as ground truth for the automatic generation of venue personality values. The next section presents related work in the fields of personality research, place recommendation and opinion mining. Section 3 presents the personality related adjectives, results of previous personality research, which are used for venue characterisation. Section 4 describes the experimental setting used to acquire data on venue personality from participants, and the results of a survey, it then presents the automated extraction approach using reviews. Finally Section 5 concludes and describes future work.

2. RELATED WORK
This research shares similar goals with opinion mining, although with different dimensions of interest. It also contributes to personality research applied to other subjects than humans, and contributes to the field of place recommendation.

Opinion Mining is concerned with applying computational methods for the detection and measurement of *opinion*, *sentiment* and *subjectivity* in text [3]. Particularly, sentiment analysis is concerned with the automated detection of negative or positive sentiment in natural text, while affective computing is concerned with the detection of human emotions such as fear, anger or humour. In both cases, the dimensions of interest are different from the ones studied in this research. Moreover, sentiment analysis applied to venues through reviews may be seen as redundant, since the review score already conveys a positive or negative judgment. This research may therefore be considered as an extension of sentiment analysis to additional dimensions of psychological interest.

Beside humans, personality traits research has been applied to non-human animals [4]. To the best of our knowledge, this constitutes the only extension of personality traits research beyond human personality, with no attempts towards other realms such as objects or places. The research presented here applies personality research to geographic entities. If extending personality traits theory to non-living structures may be controversial, as it could be argued it only applies in a metaphorical way, it also constitutes a new field of investigation potentially useful to future place recommender systems.

Tourism recommender systems have been developed to suggest venues personalized according to the user querying the system. Such venues can be generic points of interest, restaurants/cafes, or hotels (see [5] for a survey). However, the criteria used for recommendation can be described as *external*, such as *price, cuisine,* and *look*, as opposed to personality related recommendation. The present work can use the personality of a user (a *tourist*) to recommend the most adapted venue to his or her psychological profile. Beyond tourism, with the advent of mobile devices and location based services that can monitor our geographical positions in real time or at regular check points ('checkins'), mobile and online services related to places are even more popular, evolving from an initial adaptation of online maps and navigators towards services more oriented to provide reviews and personalised recommendations such as Yelp[1] and Qype[2], to others that combine location and user mobility with a social networking component, for example Foursquare[3], Flickr[4], and Google+ Local[5]. These services have evolved towards a place representation that is more related to the individual needs of users, regardless of his tourism or business needs, with users being at a particular location at a particular time often making use of tags, annotations and other user generated content which in turn informs recommendation [2]. User generated content, in the form of reviews, is used in this research to extract the personality of places.

3. KEYWORD EXTRACTION

We follow a 'bag of words' approach in which both venues and personality dimensions are represented by unigrams (individual words). Unigrams describing places are extracted from online sources providing reviews for those places, while keywords (adjectives) representing dimensions of personality are gathered from psychology literature [6]. The next sections describe the extraction and organisation of the data.

3.1 Venues Keywords

Reviews of venues are extracted from various location services using their respective APIs, or through text 'scrapping' techniques. The sources used for review extraction are: *Foursquare*, *Google Places*, *Yelp*, *Qype* and *Yell*[6].

Review text was stripped of punctuation and stopwords, uncapitalised, and part of speech (POS) tagged to filter verbs and adverbs. A negation detection algorithm is applied to the list of words to tag adjectives used in a negative sense. The algorithm is a modified version of the one presented in [3], itself inspired by NegEx [7]. The algorithm was modified to ignore so called *pseudo negations* (such as introduced by 'no increase', 'no wonder', 'no change', 'not cause', 'not only', 'not necessarily') because they were resulting in too many false positive in the reviews corpus. Also 'nothing' was added to the *prenegation* set used by the algorithm, in order to identify sentences such as "there's nothing special or exciting about…" as negatives.

3.2 Inventory of Personality Adjectives

Here we describe the inventory of personality adjectives presented by Saucier and Goldberg in [6]. The final 435-adjective inventory was obtained in their work by annotating personality related adjectives with familiarity ratings, in order to filter out the least familiar English trait personality descriptors. The resulting list of familiar adjectives was then used in that study to characterise the personality of a large sample of individuals (N=899). Following principal component analysis Saucier and Goldberg identified five clusters corresponding to five personality dimensions. Correlation scores for a selection of adjectives are given in Table 1.

Table 1 Examples of personality adjectives with FFM correlation values.

Adjective	II	I	III	IV-	V
excitable	0.22	0.3	-0.07	0.31	-0.1
friendly	0.37	0.39	-0.03	-0.17	-0.16
generous	0.4	0.15	-0.03	-0.15	0.04
independent	-0.14	0.2	0.18	-0.2	0.3
kind	0.6	0.07	0.06	0.02	0
playful	0.2	0.41	-0.12	-0.02	-0.09
quiet	0.15	-0.64	0.15	-0.09	0.12
reasonable	0.38	-0.06	0.25	-0.25	0.12
relaxed	0.21	0.17	0.04	-0.48	-0.02
traditional	0.14	-0.14	0.28	0.02	-0.36

The mapping between dimensions I to V and the OCEAN traits characterisation is given in Table 2. IV is usually inversed to represent neuroticism (IV-), i.e. high on IV- corresponds to low emotional stability. Following this correspondence, the adjective *reasonable* strongly correlates (0.38) with the OCEAN A dimension (Agreeableness) as well as with C (Conscientiousness) at 0.25, while *friendly* also correlates with A (0.37) but even more with E (Extraversion) with 0.39 as well as negatively (-0.17) with N (Neuroticism) and Openness (-0.16).

[1] http://www.yelp.co.uk/

[2] http://www.qype.co.uk/

[3] https://foursquare.com/

[4] http://www.flickr.com/

[5] http://www.google.com/+/learnmore/local/

[6] http://www.yell.com/

Table 2 Familiar adjectives dimensions to OCEAN correspondence.

I	Extraversion
II	Agreeableness/Benevolence
III	Conscientiousness
IV	Emotional Stability (neuroticism)
V	Intellect/Imagination (openness)

4. MATCHING APPROACH

To assign dimensions of personality to a venue, terms obtained from the review extraction that we performed were matched to the list of familiar personality adjectives. The ground truth was obtained from a human experiment which task consisted in selecting adjectives for a venue, from a subset of the Saucier and Goldberg familiar personality adjectives that was relevant to the type of the venue. In the following sections we describe the human experiment followed by the automated matching approach.

4.1 Human Experiment

The experiment consisted of an annotation task where participants were asked to qualify known venues with adjectives. The task consisted of two steps. In a first step, for each venue, the participants were asked to state whether or not they knew the venue and have visited it. In a second step participants were asked to select adjectives relevant to the venues they knew, from a presented set. The following sub-sections describe the places chosen, the presented set of adjectives, the experimental setting, the matching approach and results.

4.1.1 Place Selection

12 venues were selected in Cardiff, Wales, UK according to the following criteria:

- *Review availability*: selected venues had reviews from at least two Web sources,
- *Global diversity*: the places were divided into 4 categories, or types, with *T = {cafe, club, pub, restaurant}* and 3 venues of each type,
- *Category diversity*: in each category not all venues had the same characteristics,
- *Expert knowledge*: we selected venues that were most likely known to the participants.

Table 3 lists the chosen venues, accompanied with a brief description.

Table 3 Venues selected for the experiment.

venue name	venue type	Description
A shot in the dark	cafe	Independent coffee place
Ernest Willow	pub	Traditional pub
Starbuck Queen St	cafe	Chain coffe shop
Pen and Wig	pub	Traditional pub, with live music and students
Promised Land	pub	Traditional pub, with live music
Glam	club	Independent club
Oceana	club	Chain club
Pulse	club	Gay club
Greggs	restaurant	Chain bakery
Venus Kebabs	restaurant	Kebab place
Costa Coffee	cafe	Chain coffee shop
Balti King	restaurant	Indian restaurant

4.1.2 Adjectives Selection

In order to limit the human effort required to complete the task, the list of 435 familiar personality adjectives was filtered according to common collocations extracted from the Google books corpus available from *http://byu.edu*. Collocations were mined for adjectives preceding a type name, i.e. one of *{cafe, club, pub, restaurant}*. Then the resulting list of adjectives was matched to the 435 familiar adjectives.

For reasons of sense ambiguity some lists of correlations were larger than others, with keywords not immediately relevant to geographic venues. For example the type *club* had adjectives such as *charitable* or *dominant*, not immediately related to a dancing club or a nightclub, but probably more to a club as a group of individuals. Some participants complained about these terms as well as for the lack of negative terms in some cases. However, no manual filtering was done on the matched adjectives in order to preserve the objective nature of the task. Table 4 presents, for each type, the selected subset of adjectives, as well as descriptive statistics regarding the correlations of the set of adjectives to personality dimensions.

In Table 4, the minimum and maximum correlation for every dimension as well as the mean indicate the range of choice offered to the participant for a venue of a given type. In a few cases, such as for the category *pub* the range on dimension II (Agreeableness) does not allow for negative values. This is a limitation of the experiment, which may however reflect the fact that most venues tend to favour agreeableness, and that in reviews and other place descriptions, from which collocations are extracted, agreeableness terms are overrepresented.

4.1.3 Experimental setting

21 anonymous participants with knowledge of Cardiff completed the survey. Each venue was known by an average of 11.75 participants, each participant knowing an average of 6.71 venues. To address the small number of participants for some venues we used Wilson's statistical score interval to estimate the number of checks for an adjective with 85% probability, and selected only adjectives checked by at least 20% of participants.

Given the adjectives selected, OCEAN dimensions for venues are calculated first by averaging the correlations of the adjectives selected by more than 20% of users, for a venue, then by multiplying the resulting vector by 5, corresponding to the desired range of each dimension ([0,5]), and finally by adding 2.5 (half of the range) to the result.

Formally, for every adjective $a_i \in A_t$ selected for a venue v from the type's subset of adjectives A_t, the number of selected adjectives n_a, the all-ones vector, and the corresponding correlation vector $\boldsymbol{c}_i$ as given for each adjective in [6], the resulting OCEAN survey vector $\boldsymbol{os}_v$ is obtained following to Equation 1.

$$\boldsymbol{os}_v = \left(\frac{\sum_1^{n_a} \boldsymbol{c}_i}{\mathrm{n}} * 5\right) + (\mathbf{1} * 2.5)$$

Equation 1 Ocean score vector for survey results.

Table 4 Selected subsets of adjectives for venue category

Venue Type	Adjectives	n	D	D'	min corr.	max corr.	mean	variance	skewness	kurtosis
cafe	warm, pleasant, cheerful, modest, humble, friendly, quiet, sophisticated, informal, casual, intellectual, smart, bright, traditional, pretentious	15	II	A	-0.15	0.56	**0.20**	0.04	0.12	-0.93
			I	E	-0.64	0.39	0.03	0.07	-0.96	0.75
			III	C	-0.26	0.30	0.08	0.02	-0.64	-0.14
			IV-	N	-0.33	0.10	-0.11	0.02	0.26	-0.91
			V	O	-0.36	0.50	0.07	0.07	0.33	-0.80
club	cooperative, charitable, agreeable, pleasant, loyal, cheerful, modest, jovial, moral, religious, natural, rude, rough, abrupt, tough, social, enthusiastic, energetic, dominant, merry, active, friendly, competitive, enterprising, quiet, serious, efficient, formal, ambitious, conservative, sophisticated, discreet, defensive, masculine, informal, casual, intellectual, smart, philosophical, independent, progressive, diplomatic, artistic, curious, simple, conventional, traditional	47	II	A	-0.50	0.52	0.10	0.07	-0.64	-0.22
			I	E	-0.64	0.58	0.09	0.06	-0.25	0.35
			III	C	-0.26	0.57	0.12	0.03	0.17	0.26
			IV-	N	-0.43	0.41	-0.08	0.02	0.55	2.23
			V	O	-0.45	0.50	0.05	0.04	-0.06	0.28
pub	warm, pleasant, friendly, hearty, quiet, informal, casual, simple, traditional	9	II	A	0.07	0.56	**0.23**	0.03	0.88	-0.86
			I	E	-0.64	0.39	0.04	0.10	-1.05	0.40
			III	C	-0.26	0.28	0.02	0.03	-0.20	-0.36
			IV-	N	-0.33	0.10	-0.15	0.02	0.65	-0.64
			V	O	-0.45	0.23	-0.07	0.05	-0.57	-0.77
restaurant	warm, cooperative, agreeable, pleasant, cheerful, reasonable, modest, humble, boisterous, friendly, competitive, silent, quiet, sedate, serious, efficient, reliable, formal, ambitious, sophisticated, refined, discreet, informal, unassuming, casual, smart, innovative, bright, creative, independent, simple, conventional, traditional, pretentious	34	II	A	-0.25	0.56	**0.17**	0.04	0.12	-0.51
			I	E	-0.66	0.42	-0.01	0.07	-0.54	0.14
			III	C	-0.26	0.57	0.14	0.03	0.07	-0.17
			IV-	N	-0.33	0.14	-0.09	0.01	0.16	-0.82
			V	O	-0.45	0.49	0.04	0.05	-0.05	-0.12

4.1.4 Survey results

The results of the survey are presented in **Error! Reference source not found.**, with OCEAN results, number of participants knowing the venue, and average number of adjectives used.

ID	venue name	venue type	Nb of responses	Avg nb adjectives	Total adjectives	O	C	E	A	N
V01	A shot in the dark	cafe	17	5.93	15	3.05	2.69	2.81	3.65	1.89
V02	Ernest Willow	pub	17	5.00	9	1.78	2.34	2.28	3.06	1.59
V04	Starbuck Queen St	cafe	16	3.83	15	2.49	2.15	3.21	3.55	1.90
V10	Pen and Wig	pub	16	7.33	9	2.31	2.49	3.29	3.80	1.72
V13	Promised Land	pub	13	4.38	9	2.51	1.98	3.38	3.76	1.29
V21	Glam	club	9	2.24	46	2.53	2.60	4.46	2.53	2.16
V27	Oceana	club	10	2.33	47	2.36	2.45	4.13	1.83	2.05
V33	Pulse	club	10	1.95	47	2.60	3.25	4.65	3.05	1.80
V52	Greggs	restaurant	19	3.65	34	1.98	4.03	1.98	3.68	1.98
V53	Venus Kebabs	restaurant	2	1.00	34	1.68	3.18	1.98	4.00	2.15
V55	Costa Coffee	cafe	4	1.70	15	2.58	2.14	3.46	4.07	1.63
V57	Balti King	restaurant	8	1.89	34	2.38	3.07	2.98	3.35	1.57

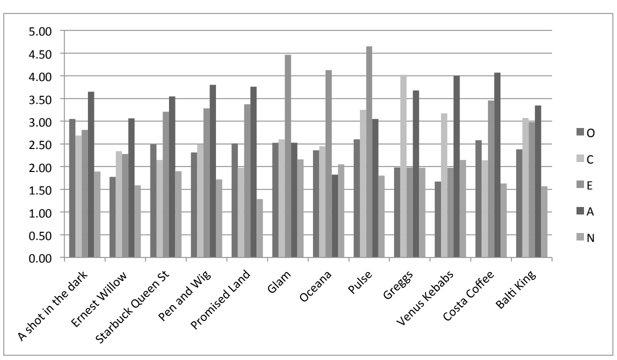

Figure 4.1 Survey results.

The following comments may be made on the results:

- V01, the independent coffee place, ranks the highest in O, due to the use of adjectives such as ‘intellectual’ and ‘sophisticated’ (0.5 and 0.18 correlation on O, respectively).
- V02, a traditional pub, ranks low on O, due to the use of adjectives such as ‘simple’ and ‘traditional’ (-0.45 and -0.36 on O, respectively)
- all *clubs* rank above 4 on E, while all other types of venues are below.
- *pubs* rank higher on E than *cafés*, but only on average, with V10 and V13, both pubs having live music, contrasting with V02, a traditional chain owned pub.
- Neuroticism is constantly below average, which suggests that venue owners don’t encourage negative emotions such as fear and anxiety.

The list of adjectives used for every venue is given in Table 5.

Table 5 Adjectives selected for every venue

ID	Adjectives selected (20% of participants at 85% confidence)
V01	warm, friendly, pleasant, casual, informal, intellectual, quiet, sophisticated
V02	simple, informal, casual, traditional
V04	informal, casual, warm, pleasant, pretentious
V10	traditional, pleasant, warm, friendly, informal, hearty, casual
V13	informal, pleasant, casual, friendly
V21	social, rude, energetic, active
V27	energetic, rude, social, rough
V33	active
V52	simple, conventional, reasonable, efficient, modest, reliable
V53	agreeable, traditional
V55	warm, informal, pleasant, friendly, casual
V57	casual, reasonable, friendly, independent, conventional

4.2 Reviews Matching

The approach taken to match familiar personality adjectives to venues differs slightly from the experiment, due to some reviews being short, and therefore not providing an exhaustive description of the personality related aspects of the venue, or a type related one. Therefore, the full list of familiar adjectives was used, rather than type correlated adjectives, and keywords were stemmed to encourage more matches.

The negation algorithm is applied to venues keywords, to obtain a negation marker (1 or -1) for each term. Venues keywords extracted from reviews as well as familiar adjectives are stemmed using the Porter stemmer. Stemmed review terms are then matched to the list of 435 stemmed familiar adjectives. In case of a match, the correlation values of the matching adjective is added, possibly negated if a negation has been detected. The resulting sum is then averaged, multiplied by the range (5), and added to the average value for a dimension (2.5).

Formally, for every adjective $a_i \in A$ found in reviews for a venue v, the number of found adjectives n, the all-ones vector, and the corresponding correlation vector $\boldsymbol{c}_i$ as established in Saucier, and the sign s_a produced by the negation detection algorithm, the resulting OCEAN review vector $\boldsymbol{or}_v$ is obtained according to Equation 2.

$$\boldsymbol{or}_v = \left(\frac{\sum_1^{\mathrm{n}} s_a \boldsymbol{c}_i}{\mathrm{n}} * 5\right) + (\mathbf{1} * 2.5)$$

Equation 2 Ocean score vector for automated personality extraction.

4.2.1 *Results of automatic matching*

Results of automated reviews matching are presented in Figure 4.2.

ID	venue name	venue type	O	C	E	A	N
V01	A shot in the dark	cafe	2.66	2.85	2.94	3.41	1.86
V02	Ernest Willow	pub	2.71	2.96	2.54	3.43	2.09
V04	Starbuck Queen St	cafe	2.39	3.26	2.89	3.71	2.09
V10	Pen and Wig	pub	2.5	3.13	2.94	3.8	1.82
V13	Promised Land	pub	2.43	2.91	2.78	3.81	2.06
V21	Glam	club	2.66	2.64	3.03	2.98	2.38
V27	Oceana	club	2.55	2.69	2.97	3.34	2.25
V33	Pulse	club	2.56	2.62	3.2	2.94	2.18
V52	Greggs	restaurant	2.94	2.97	1.73	2.43	1.67
V53	Venus Kebabs	restaurant	2.25	2.74	3.32	3.45	2.22
V55	Costa Coffee	cafe	2.77	3.08	2.87	2.77	2.05
V57	Balti King	restaurant	2.45	2.85	2.82	2.95	1.53

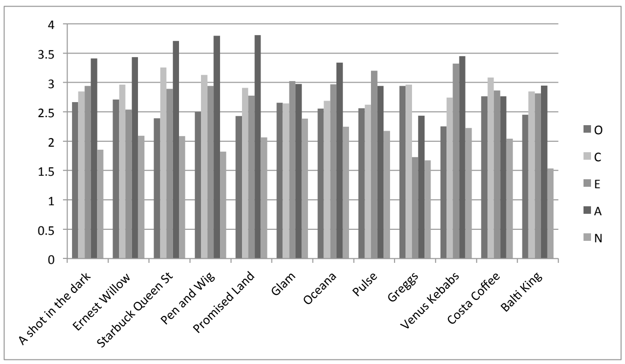

Figure 4.2 Automated matching results

The following comments may be made on the results:

- V01, the independent coffee place, is still ranking high in O, but is only 4th position. This is due to an abundance of agreeableness terms present in reviews, such as 'friendly' (stem: 'friendli'), which also affect the O dimension (-0.16 on O).
- V02 also ranks high in O due to the term 'reasonable', qualifying quality and prices, and correlating with O (0.12) and with A (0.38)
- *Clubs* rank the highest on E, although V53 ranks the highest, due to a majority of 'friendly' (0.39 on E)
- Neuroticism is constantly below average, as in the survey.

The list of stemmed adjectives used in reviews of each venue is given in Table 6.

Table 6 Ordered list of adjective stems extracted from reviews, with number of use

ID	Ordered list of stems with number of uses
V01	reason (10), relax (8), friendli (7), gener (4), kind (4), pretenti (3), sophist (3), …
V02	reason (6), gener (3), quiet (3), pleasant (2), earnest (2), excit (1), …
V04	kind (1), prompt (1), consist (1), friendli (1), pleasant (1), warm (1), cold (1)
V10	reason (13), relax (6), friendli (5), kind (4), tradit (3), help (2), competit (2), …
V13	friendli (6), tradit (3), thought (3), reason (3), excit (2), help (2), relax (2), quiet (2), …
V21	gener (3), indulg (2), activ (2), sophist (2), excit (1), frivol (1), …
V27	reason (8), play (7), talk (7), thought (5), cold (3), help (3), depend (3), suggest (3), …
V33	friendli (4), merri (2), versatil (2), quiet (2), reason (2), opinion (2), excit (1), …
V52	excit (2), imagin (1), independ (1), quiet (1), invent (1), bland (1), principl (1), simpl (1)
V53	friendli (6), cold (2), effici (1), help (1), thought (1), fear (1), talk (1)
V55	independ (2), kind (1), depend (1), rude (1), relax (1),…
V57	reason (1), unsympathet (1), friendli (1)

4.2.2 *Evaluation and Discussion*

To characterise the personality of a venue in a human-friendly way, the OCEAN acronym was used with the characters reordered according to dimension values, in decreasing order. For example an open and extrovert place can have 'OEACN' as profile, compared to an agreeable but not very open place which would be an 'ACENO'. This *ocean signature* was used to compare results from the survey to results of the automated process. The two signature strings were compared using the Hamming string distance, which counts matching characters in two strings of equal length.

Although the average distance is 3.9 (out of a maximum of 5), the first dimension matches 8 times out of 12, suggesting that the automated method provides good results to determine the most important characteristic of a venue.

For every venue, we also calculated a mean error using Equation 3:

$$me_v = \frac{\sqrt{(\boldsymbol{or_v} - \boldsymbol{os_v})^2}}{5}$$

Equation 3 mean error for a venue

When averaging the error across the sample, we obtained a value of 0.81, with a maximum value of 1.04, and a minimum value of 0.6, suggesting that, on average, the algorithm matches human evaluation with an error lower than 20%.

Some of the imprecision found in comparing survey and automated matching results are certainly due to the difficulty of accurately parsing reviews. For example, Greggs (V52), a chain owned bakery, appears in the automated process as an 'exciting', 'imaginative' and 'independent' place due to reviews sentences such as:

- "I think there are a lot of other places people can go that serve up *imaginative* food",
- "but ultimately, there's nothing special, or *exciting*, or especially tasty in a Greggs"
- "At the same time, Greggs […] cannot compete with excellent fresh, local, *independent* bakeries".

In all these sentences the negative use of an adjective has been wrongly interpreted as a positive by the particular variant of the *NegEx* algorithm used.

5. CONCLUSIONS AND FUTURE WORK

This work constitutes a first attempt at a qualification of venues according to their personality using an automated matching method between reviews and personality related adjectives. It provides both an experimental approach of venue personality annotation by human subjects, as well as an automated venue personality extraction approach based on reviews.

We demonstrated that reviews provide a reliable source of personality related adjectives, matching human evaluation in several aspects, notably the main personality dimension of a venue and the range of personality values. Further work includes extending and modifying the evaluation approach, a better handling of negation, and possibly an evaluation using personality tests of users of the venue.

A larger scale evaluation of the automated approach is needed, requiring more human experiment samples. The principal difficulty for such an evaluation is that participants are required to have knowledge of the venue. A possibility to explore is to distribute questionnaires in a venue, which would ensure the punters already know the place. This outreach effort could be combined with the administration of a FFM personality test for every respondent. The adjective list for every venue could be adapted by using negatives or a Likert scale for each, which will ensure a wider range of personality is available for a type of venue.

Further work on the improvement of the negation detection algorithm would increase the quality of the automated results. However it is unclear whether negation detection algorithms exist which could accurately address the examples presented above.

6. ACKNOWLEDGMENTS

This research has been funded by RECOGNITION grant 257756, an EC - FP7 Future Emerging Technologies project concerning Self-Awareness in Autonomic Systems. Many thanks to Steven Schockaert for comments and pointers that helped informing some of the methods described in this article.

7. REFERENCES

[1] Y.-F. Tuan, *Space and Place: The Perspective of Experience*, 8(null) ed. University of Minnesota Press, 2001.

[2] R. Purves, A. Edwardes, and J. Wood, “Describing place through user generated content,” *First Monday*, 2011. [Online]. Available: http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/3710. [Accessed: 25-Jun-2012].

[3] B. Ohana, “Opinion mining with the SentWordNet lexical resource,” Dublin Institute of Technology, 2009.

[4] T. A. R. Weinstein, J. P. capitanio, and S. D. Gosling, “Personality in animals,” in *Handbook of Personality: Theory and Research*, no. 12, O. P. John, R. W. Robins, and L. A. Pervin, Eds. 2008.

[5] Katerina Kabassi, “Personalizing recommendations for tourists,” *Telematics and Informatics*, vol. 27, no. 1, pp. 51–66, Feb. 2010.

[6] G. Saucier and L. R. Goldberg, “Evidence for the Big Five in analyses of familiar English personality adjectives,” *Eur. J. Pers.*, vol. 10, no. 1, pp. 61–77, 1996.

[7] W. W. Chapman, W. Bridewell, P. Hanbury, G. F. Cooper, and B. G. Buchanan, “A Simple Algorithm for Identifying Negated Findings and Diseases in Discharge Summaries,” *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 34, no. 5, pp. 301–310, Oct. 2001.