Linda Cappellato and Nicola Ferro and Gareth Jones and Eric San Juan (eds.):
CLEF 2015 Labs and Workshops, Notebook Papers, 8-11 September, Toulouse, France.
CEUR Workshop Proceedings. ISSN 1613-0073, http://ceur-ws.org/Vol-1391/, 2015.

Overview of the 3rd Author Profiling Task at PAN 2015

Francisco Rangel[1,2] Fabio Celli[3] Paolo Rosso[1]
Martin Potthast[4] Benno Stein[4] Walter Daelemans[5]

[1]Natural Language Engineering Lab, Universitat Politècnica de València, Spain
[2]Autoritas Consulting, S.A., Spain
[3]University of Trento, Italy
[4]Web Technology & Information Systems, Bauhaus-Universität Weimar, Germany
[5]CLiPS - Computational Linguistics Group, University of Antwerp, Belgium

pan@webis.de http://pan.webis.de

Abstract We overview the framework and the results for the Author Profiling Shared Task organised at PAN 2015. This year's task aims at identifying age, gender, and personality traits of Twitter users. With the help of an online personality test[1] a dataset was collected from Twitter with annotations about age, gender, and the Big Five personality traits in the four languages English, Spanish, Italian, and Dutch. The paper in hand presents the approaches of 22 participants.

1 Introduction

Author profiling distinguishes between classes of authors by studying their sociolect aspect, i.e., how language is shared or how an author can be characterized from a psychological viewpoint. This information helps in identifying profiling aspects such as gender, age, native language, or personality type. Author profiling is a problem of growing importance, among others for applications in forensics, security, and marketing. From a forensic linguistics perspective, for example, one would like to learn about the linguistic profile of the author of a harassing text message (language used by a certain type of people) and identify certain characteristics (language as evidence). From a marketing viewpoint, companies may be interested to learn about the demographics of people who like or dislike their products, given blogs and online product reviews as analysis source.

In the Author Profiling Shared Task at PAN 2013[2], the identification of age and gender relied on a large corpus collected from social media [64]. At PAN 2014[3], the objective was to extend the age and gender identification task to new genres, namely social media, blogs, Twitter, and hotel reviews [63]. Except for the hotel reviews subcorpus used at PAN 2014, which was available in English only, all documents were

[1] In order to address ethical and privacy issues, authors were asked for their permission to use the tweets when answering the personality test. The dataset was anonymised, password protected, and released to task participants only.
[2] http://webis.de/research/events/pan-13/pan13-web/author-profiling.html
[3] http://webis.de/research/events/pan-14/pan14-web/author-profiling.html

provided in both English and Spanish. This year, besides the focus on age and gender identification, we introduce the task of personality recognition. Furthermore, contrary to most of the existing research in computational linguistics [3] and social psychology [53], which focuses on the English language, we broaden the analysis in terms of languages: in addition to English, we provide data also in Spanish, Italian, and Dutch.

The remainder of this paper is organised as follows. Section 2 covers the state of the art, Section 3 describes the corpus and the employed evaluation measures, and Section 4 presents the approaches submitted by the participants. Section 5 and 6 discuss results and draw conclusions respectively.

2 Related Work

The following sections describe the related work for age and gender identification as well as personality traits recognition.

2.1 Age and Gender Identification

Pennebaker's [54] investigated how the style of writing is associated with personal attributes such as age and gender. In [3] the authors approached the task of gender identification by combining function words with parts-of-speech (POS). They analysed formally written texts extracted from the British National Corpus and achieved approximately 80% classification accuracy. Similar research was done by the authors in [26, 8].

Recently, most investigations focus on social media. In [31], for example, the authors studied the problem of automatically determining an author's gender by proposing combinations of simple lexical and syntactic features; they achieved an accuracy of about 80%. Schler *et al.* [68] studied the effect of age and gender in the style of writing in blogs; the authors gathered over 71,000 blogs and obtained a set of stylistic features like non-dictionary words, parts-of-speech, function words and hyperlinks, combined with content features, such as word unigrams with the highest information gain. They obtained an accuracy of about 80% for gender identification and about 75% for age identification.

We want to point out that the previously described studies were conducted with texts of a length of at least of 250 words. The effect of data size is known, however, to be an important factor in machine learning algorithms. Zhang and Zhang [77] experimented with short segments of blog post, specifically 10,000 segments with 15 tokens per segment, and obtained 72.1% accuracy for gender prediction, compared to more than 80% in the previous studies. Similarly, Nguyen *et al.* [46] studied the use of language and age among Dutch Twitter users, where the documents are really short, with an average length of less than 10 terms. They modelled age as a continuous variable as they had previously done in [47] and used an approach based on logistic regression. They also measured the effect of gender on the performance of age detection, considering both variables as inter-dependent, and achieved correlations of up to 0.74 and mean absolute errors between 4.1 and 6.8 years.

With regard to the shared task on Author Profiling at PAN [64, 63], most participants used combinations of style-based features such as frequency of punctuation marks, capital letters, quotations, together with POS tags and content-based features such as Latent

Semantic Analysis, bag-of-words, TF-IDF, dictionary-based words, topic-based words. Irrespective of the good performance of n-gram features as reported in [27] and [52], it is worth to mention the effect of more elaborated features as well. For example, the second order representations based on relationships between documents and profiles by the best performing team at PAN-AP 2013 and 2014 [35, 34], or the use of collocations by the best performing team on English data in 2013 [40]. Recently, The EmoGraph [62] graph-based approach tries to capture how users convey verbal emotions in the morphosyntactic structure of the discourse, obtaining competitive results with the best performing systems at PAN 2013. Moreover, using the PAN-AP-2013 dataset, the authors in [75] investigate a high variety of different features and show the contribution of information-retrieval-based features in age and gender identification. In [36], the authors approach the task with 3 million features in a MapReduce configuration, obtaining high accuracies with fractions of processing time.

2.2 Personality Recognition

Personality may be defined along five traits using the Five Factor Theory [15], which is the most widely accepted in psychology. The five traits are: *extroversion* (E), *emotional stability / neuroticism* (S), *agreeableness* (A), *conscientiousness* (C), and *openness to experience* (O). The automatic recognition of personality from text has been addressed by pioneering works about 10 years ago. Argamon *et al.* [72] focused on two of the Big Five traits (Extraversion and Emotional Stability), measured by means of self-reports. They used Support Vector Machines, trained on word categories and relative frequency of function words, to recognize these two traits. In a similar way, Oberlander and Nowson [49] worked on the classification of personality types of bloggers by extracting patterns in a bottom-up fashion. Mairesse *et al.* [38], investigated systematically the usefulness of different sets of textual features exploiting psycholinguistic dictionaries (LIWC[4] and MRC[5]). They extracted personality models from self-reports and observed data, and they reported that the openness to experience trait yield the best performance.

In more recent years, the interest in personality recognition has developed into two areas: the analysis of human behaviour and social network analysis. Several studies have started exploring the wealth of behavioral data made available by cameras, microphones [56, 6, 43, 33, 2], wearable sensors [50, 28], and mobile phones [70, 14, 44], by linking personality traits to dimensions such as face to face interaction or speech, video, and text transcriptions. But, researchers have also focused on personality prediction using corpora of social network data, such as Twitter and Facebook, exploiting either linguistic features in status updates, social features such as friends count, and daily activity [19, 60, 13]. Kosinski *et al.* [32] made an extensive analysis of different features, including the size of friendship network, the uploaded photos count and attended events, finding the correlations with the personality traits of 180,000 Facebook users. They reported very good results in the automatic prediction of Extraversion. Bachrach *et al.* made an extensive analysis of the network traits (size of friendship network, uploaded photos, events attended, the frequency a user has been tagged in photos)

[4] http://www.liwc.net/
[5] http://www.psych.rl.ac.uk/

that correlate with personality of 180,000 Facebook users. They predicted personality scores using multivariate linear regression, and reported good results on extroversion. Schwartz *et al.* [69] analyzed 700 million words, phrases, and topic instances collected from the Facebook messages of 75,000 volunteers, who also filled a standard Big Five personality test. Their results showed interesting correlations among words usage and personality traits. For example, extroverts were more likely to mention social words, whereas introverts were more likely to mention words related to solitary activities. Neurotic people, however, disproportionately use the phrases "sick of" and the word "depressed", whereas emotionally stable individuals wrote about enjoyable social activities that may foster greater emotional stability, such as "sports", "vacation", "beach", "church", or "team". Since almost all researchers who worked in personality recognition used different evaluation measures and procedures, it is not easy to exactly define the state-of-the-art regarding the classification effectiveness. This fact lead to evaluation campaigns such as the Workshop on Computational Personality Recognition [12].

There are many applications of automated personality recognition, for example in security and deception detection [16], or recommendation systems [66]. Recently, it has been found that computer-based personality judgments are more accurate than those made by humans [76]. A very good overview of the recent work done in automatic personality recognition can be found in Vinciarelli and Mohammadi [73].

3 Evaluation Framework

This outlines the construction of the corpus, highlighting particular properties, challenges, and novelties. Moreover, the evaluation measures and the software submission procedure based on TIRA are described.

3.1 Corpus

We have collected the PAN-AP-2015 corpus from Twitter in the four languages English, Spanish, Italian, and Dutch. The corpus is annotated with gender and personality traits as well as with age classes (English and Spanish only). The age and gender information was reported by the Twitter users themselves. For labelling age, the following classes were considered: *a)* 18-24, *b)* 25-34, *c)* 35-49, and *d)* 50+. Personality traits were self-assessed with the BFI-10 online test [61] and reported as scores normalized between -0.5 and +0.5 (the mean for each trait is reported in Table 1). As in the previous edition of this task, the dataset was split into three parts, namely for training, for early birds, and for test. The distribution of the labels in the corpus is reported in Figure 1. The corpus is balanced wrt. gender, but the skew of the age distribution is considerable due to the lower number of aged 50 and older using Twitter.[6]

3.2 Performance Measures

The evaluation of the participants' approaches relies on two different measures. For age and gender identification the accuracy measure was used. In particular, we calculated the ratio between the number of correctly predicted authors by the total number

[6] http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/

Table 1. Distribution of Twitter users with respect to the labels in the corpus per language.

	Training				Early birds				Test			
	EN	ES	IT	DU	EN	ES	IT	DU	EN	ES	IT	DU
Users	152	110	38	34	42	30	12	10	142	88	36	32
18-24	58	22			16	6			56	18		
25-34	60	56			16	14			58	44		
35-49	22	22			6	6			20	18		
50+	12	10			4	4			8	8		
Male	76	55	19	17	21	15	6	5	71	44	18	16
Female	76	55	19	17	21	15	6	5	71	44	18	16
E (mean)	0.16	0.18	0.17	0.24	0.19	0.15	0.16	0.21	0.17	0.16	0.15	0.24
S (mean)	0.14	0.07	0.20	0.21	0.11	0.07	0.24	0.23	0.13	0.09	0.20	0.22
A (mean)	0.12	0.14	0.22	0.13	0.14	0.17	0.17	0.14	0.14	0.14	0.19	0.15
C (mean)	0.17	0.24	0.18	0.14	0.17	0.22	0.22	0.17	0.17	0.21	0.21	0.17
O (mean)	0.24	0.18	0.23	0.29	0.28	0.19	0.29	0.27	0.26	0.19	0.25	0.28

of authors. We calculated individual accuracy scores for each language, gender, and age class. Finally, we combined the accuracy values to obtain a joint identification of age and gender in each language. For personality recognition the Root Mean Square Error (RMSE) for each trait in each language was computed using Equation 1, where n denotes the number of authors, f_i the value for trait i, and $\widehat{f_i}$ the predicted one.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\widehat{f_i} - f_i)^2}{n}} \tag{1}$$

RMSE measures the distance of the predicted value to the true value for this trait. It is a measure of error, so the lower is the score, the better is the performance. The overall RMSE per language was computed as the arithmetic mean of each trait RMSE. We have combined the joint accuracy with the global RMSE following Formula 2.

$$rank = \frac{(1 - RMSE) + joint\_accuracy}{2} \tag{2}$$

Finally, the global rank is calculated as the arithmetic mean of the previous ranks.

3.3 Software Submissions

We continued to invite software submissions instead of run submissions for the third time. Using software submissions, participants are asked to submit executables of their author profiling software instead of just the output (also called "runs") of their softwares on a given test set. Our rationale to do so is to increase the sustainability of our shared task and to allow for the re-evaluation of approaches to Author Profiling later on, for example, on future evaluation corpora. To facilitate software submissions, we developed the TIRA experimentation platform [21, 22], which provides service to handle software submissions as simple as run submissions. Using TIRA, participants deploy their software on virtual machines at our site, which allows us to keep them in a running state [20].

4 Overview of the Submitted Approaches

This year 22 teams have submitted software and notebook papers. On the basis of what they explained in their notebook papers, this sections presents a summary of their approaches in terms of preprocessing steps, features, and classification algorithms.

Preprocessing Most participants carried out some kind of preprocessing, however, in first place to remove HTML code from the tweets [4, 24, 45, 65]. Also hashtags, urls and mentions were handled by many participants [4, 23, 24, 37, 48]. For example in [23] the authors changed mentions, urls, and hashtags for predefined tokens. Similarly, in [37] the authors replaced urls with the URL token, or in [37] the urls were completely removed. Although the dataset was cleaned before releasing, in [5, 58] the authors preprocessed tweets to remove RTs and shares. In [7] the authors lowercased the texts and removed numbers and stop words, such as in [74] where the authors also applied stemming. The authors in [48] removed all character sequences representing emojis in the original tweets, and the authors in [57] removed tweets with fewer than five words. Regarding feature selection, the authors in [41] applied Support Vector Machines (SVM), Recursive Feature Extraction (RFE) [25], and Fordward-Backward for age/gender and personality traits.

Features Similar to the previous editions, participants approached the task with combinations of style-based and content-based features, as well as their combination in n-gram models [7, 29, 58]. For example, character n-grams [23, 37, 71], word n-grams [4, 18, 45, 48, 51], TF-IDF n-grams [7, 18, 24, 51, 71], POS n-grams [23, 51]. In [57], the authors obtained 10 different kinds of n-grams (lemmas, words, relations, POS, etc.) from the dependency tree.

Participants combined also a high variety of different stylistic features, such as punctuation signs [7, 41, 51, 55], emoticons [7, 48, 51, 57, 59], word length [24], sentence length [55], character flooding [18, 29, 48], verbosity [71], letter case [18, 24, 29], question marks [37], and question sentences [55]. Other participants took advantage of specific Twitter elements, such as links, hashtags, or mentions [18, 24, 29, 41, 48, 51, 57].

Regarding the content-based features, the authors in [37, 39, 41, 58, 65] used topic modelling with Latent Semantic Analysis (LDA). More shallow features were used in [30] where authors obtained the 200 most frequent terms, or the combination of bag-of-words with other features in [51]. Moreover, in [37] the authors use Family Tokens (my wife/husband, my girlfriend/boyfriend, my hubby, my bf, etc.), and in [45] the authors use the most discriminant words among classes. Finally, Named Entities were taken into account in [48].

Resources such as LIWC [4, 5, 41] or NRC [18, 29] where considered to obtain psycholinguistic features such as polarity words and emotions [18, 29, 48, 51, 59]. Also other dictionaries, some of them were manually compiled, with Ironic Words [51], Taboo Words [51], or Informative Words [5, 7] where employed to evaluate specific properties of the contents.

Specific features were used in [74], where participants obtained features employed in information retrieval (IR) such as the cosine similarity or the Okapi BM25 document model. Finally, we would highlight the best performing team [1], who combined Latent Semantic Analysis (LSA) with second order features based on relationships among terms, documents, profiles, and sub-profiles.

Classification Approaches All participants approached the tasks as a machine learning problem, i.e., a classification problem to predict age and gender, and a classification or regression problem to predict personality traits. Most participants employed Support Vector Machines. For example, in [1, 23, 45, 57] all tasks were approached with LibLinear. In [7, 24, 29, 48, 58] SVMs were used for the classification of age and gender, and regression for personality recognition. Other participants used decision trees, such as the authors in [5], who used SVM for classification but Random Forest for regression, or the authors in [51, 65], who used Random Forest and J48 respectively for classification and regression. Moreover, the authors in [39] used Rotation Forests for both age and gender as well as personality traits. With regard to other algorithms, Bagging was used for regression in [39], Linear Discriminant Analysis for regression in [41], Stochastic Gradient Descent for classification and Ensemble of Regressor Chains Corrected for regression in [4], Linear Regression in [18], Ridge for regression in [71], Logistic Regression in [37], and distance-based approaches in [30, 55, 59].

5 Evaluation and Discussion of the Submitted Approaches

In this section we show a summary of the obtained results for the 22 teams. Table 2 comprises the ranking and the overall performance per language. Observe that the highest accuracies were obtained on the Dutch dataset, with values over 90% in some cases. Note that this dataset contains the smallest number of authors. The worst results were obtained in the English dataset, which also contains the highest number of authors. However, the results may be explained by the absence of age identification in Dutch, which renders the task easier. Similar effects can be observed with more related languages such as Italian and Spanish, where accuracies for the first one are higher.

The approach of *alvarezcarmona15* [1] performs overall best, and it is among the top three in every language. The authors combined the Second Order Representation, which allowed them to obtain the best results in PAN task in 2013 [35] and 2014 [34], along with LSA. On the other hand, *gonzalesgallardo15* [23] and *grivas15* [24] achieved results very close to *alvarezcarmona15*, and they are also among the top three in every language. The team of *gonzalesgallardo15* used combinations of char and POS n-grams, and *grivas15* combined TF-IDF n-grams with style-based features.

We carried out Student's t-test, which shows no significant difference between *alvarezcarmona15* and *gonzalesgallardo15* ($z_{0.05} = 0.1918 < 1.960$), *grivas15* ($z_{0.05} = 1.0444 < 1.960$) and *kocher15* ($z_{0.05} = 1.6588 < 1.960$) at 95% of confidence, and between *alvarezcarmona15* and *sulea15* ($z_{0.01} = 2.0073 < 2.3260$) at 99%. The teams of *ashraf15*, *kiprov15* and *markov15* participated not in all languages. We would like to

Table 2. Global ranking as average of each language global accuracy.

Ranking	Team	Global	English	Spanish	Italian	Dutch
1	alvarezcarmona15	**0.8404**	**0.7906**	**0.8215**	0.8089	**0.9406**
2	gonzalesgallardo15	0.8346	0.7740	0.7745	**0.8658**	0.9242
3	grivas15	0.8078	0.7487	0.7471	0.8295	0.9058
4	kocher15	0.7875	0.7037	0.7735	0.8260	0.8469
5	sulea15	0.7755	0.7378	0.7496	0.7509	0.8637
6	miculicich15	0.7584	0.7115	0.7302	0.7442	0.8475
7	nowson15	0.7338	0.6039	0.6644	0.8270	0.8399
8	weren15	0.7223	0.6856	0.7449	0.7051	0.7536
9	poulston15	0.7130	0.6743	0.6918	0.8061	0.6796
10	maharjan15	0.7061	0.6623	0.6547	0.7411	0.7662
11	mccollister15	0.6960	0.6746	0.5727	0.7015	0.8353
12	arroju15	0.6875	0.6996	0.6535	0.7126	0.6843
13	gimenez15	0.6857	0.5917	0.6129	0.7590	0.7790
14	bartoli15	0.6809	0.6557	0.5867	0.6797	0.8016
15	ameer15	0.6685	0.6379	0.6044	0.7055	0.7260
16	cheema15	0.6495	0.6130	0.6353	0.6774	0.6723
17	teisseyre15	0.6401	0.7489	0.5049	0.6024	0.7042
18	mezaruiz15	0.6204	0.5217	0.6215	0.6682	0.6703
19	bayot15	0.6178	0.5253	0.5932	0.6644	0.6881
	ashraf15	-	0.5854	-	-	-
	kiprov15	-	0.7211	0.7889	-	-
	markov15	-	0.5890	0.5874	-	0.6798

highlight the results from *kiprov15* [29] for Spanish, who obtained the second position with large features set containing n-grams, POS tags, character flooding, average sentence length, Twitter-specific features such as hashtags, urls, mentions, re-tweets, and different dictionaries such as NRC Hashtag Emotion Lexicon [42], Bad Words Lexicon,[7] or World Well-Being Project Personality Lexicon [67].

Figure 1 shows the distribution of accuracies per language. The participants obtained accuracies between 0.5049 and 0.8215. Also, the results are more concentrated below the median (0.6547). The results in Spanish are the most sparse ones. Except for Dutch, there are slightly more outliers in the lower bound; in Dutch, the outliers occur in the upper bound, for instance with accuracies over 90%.

Table 3 shows the best results per language and task. Compared to previous editions of this shared task, the this year's approaches obtained a significantly higher accuracy for both age and gender prediction. This fact may suggest that, regardless of the shorter length of individual tweets and their informality, the number of tweets per author is sufficient to predict age and gender. With regard to personality recognition one can see that the best results were obtained for Italian and Dutch. This fact can be explained by the smaller number of authors for these languages, which both reduces the variance in the data and raises the analyzable data size per author. A similar phenomenon has been reported for personality prediction given Facebook profile pictures [11]. Again, this is

[7] A combination of manually assembled dictionary and Google's "what do you love" profanity dictionary. https://opennlp.apache.org

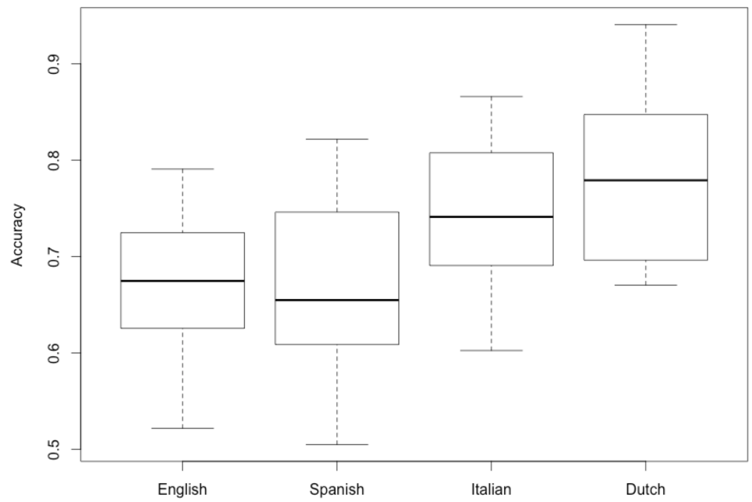

Figure 1. Distribution accuracies per language.

explained by a reduction of diversity. With regard to personality recognition, the *Stable* trait is more difficult to be predicted than the traits *Conscientious* and *Openness*.

Table 3. Best results per language and tasks

	Age and Gender			Personality Traits					
Language	*Joint*	Gender	Age	*RMSE*	E	S	A	C	O
English	0.7254	0.8592	0.8380	0.1442	0.1250	0.1951	0.1305	0.1101	0.1198
Spanish	0.7727	0.9659	0.7955	0.1235	0.1319	0.1631	0.1034	0.1017	0.1108
Italian	-	0.8611	-	0.1044	0.0726	0.1555	0.0527	0.1093	0.0972
Dutch	-	0.9688	-	0.0563	0.0750	0.0637	0.0000	0.0619	0.0354

The Tables 4 to 7 detail the results per language. Regarding the English results (Table 4), the high values for age, gender, and joint identification obtained by *alvarez-carmona15*, *gonzalesgallardo15*, and *grivas15* should be noted, all of which are among the top three in every language. With regard to RMSE, *sulea15* [71] obtained the lowest values in most traits, especially for the *Stable* trait. In order to remind the approaches, *alvarezcarmona15* used Second Order features with LSA, *gonalezgallardo15* combinations of char and POS n-grams, *grivas15* combined TF-IDF n-grams with stylistic features, and *sulea15* combined character and TF-IDF n-grams with style-based features such as verbosity ratio.

Table 4. Evaluation results in terms of accuracy for age and gender identification (left) and RMSE in personality recognition (right) on **English** texts.

	Age and Gender			Personality Traits					
Team	*Joint*	Gender	Age	*RMSE*	E	S	A	C	O
alvarezcarmona15	**0.7254**	**0.8592**	**0.8380**	**0.1442**	0.1278	0.2253	**0.1305**	0.1172	0.1202
ameer15	0.5070	0.6901	0.7183	0.2313	0.2131	0.3172	0.2154	0.1959	0.2149
arroju15	0.5704	0.7676	0.7042	0.1713	0.1636	0.2349	0.1513	0.1481	0.1584
ashraf15	0.3944	0.5563	0.6972	0.2236	0.2084	0.3151	0.1910	0.1897	0.2138
bartoli15	0.4718	0.6479	0.7465	0.1605	0.1480	0.2323	0.1360	0.1418	0.1445
bayot15	0.2465	0.5000	0.5915	0.1958	0.2137	0.2308	0.1634	0.1866	0.1844
cheema15	0.4225	0.5915	0.6690	0.1965	0.1878	0.2612	0.1766	0.1610	0.1959
gimenez15	0.3873	0.6338	0.5986	0.2039	0.1770	0.2781	0.1754	0.1819	0.2073
gonzalesgallardo15	0.6972	0.8521	0.7817	0.1491	0.1303	0.2151	0.1480	**0.1101**	0.1422
grivas15	0.6690	**0.8592**	0.7465	0.1716	0.1411	0.2039	0.1432	0.2249	0.1450
kiprov15	0.5915	0.8451	0.7254	0.1493	0.1416	0.2123	0.1411	0.1318	**0.1198**
kocher15	0.5563	0.7113	0.7113	0.1489	0.1417	0.2062	0.1427	0.1181	0.1358
maharjan15	0.5634	0.7465	0.6901	0.2388	0.2299	0.2647	0.2127	0.2222	0.2645
markov15	0.3662	0.5915	0.5845	0.1882	0.1806	0.2708	0.1570	0.1893	0.1434
mccollister15	0.5141	0.7254	0.7183	0.1649	0.1537	0.2205	0.1513	0.1443	0.1545
mezaruiz15	0.2183	0.5000	0.4085	0.1749	0.1676	0.2392	0.1572	0.1526	0.1582
miculicich15	0.5704	0.7887	0.6901	0.1475	**0.1250**	0.2247	0.1322	0.1330	0.1225
nowson15	0.3732	0.7746	0.4930	0.1655	0.1665	0.2059	0.1647	0.1483	0.1419
poulston15	0.5211	0.6901	0.7394	0.1725	0.1381	0.2223	0.1918	0.1749	0.1352
sulea15	0.6197	0.7676	0.7887	**0.1442**	0.1318	**0.1951**	0.1396	0.1297	0.1246
teisseyre15	0.6479	0.8310	0.7535	0.1500	0.1371	0.1990	0.1480	0.1309	0.1351
weren15	0.5563	0.7606	0.7042	0.1851	0.1597	0.2593	0.1768	0.1574	0.1722

Table 5 shows the results for Spanish. The high values for gender identification obtained by *alvarezcarmona15* (0.9659), *grivas15* (0.9432) and *kiprov15* (0.9091) is extraordinary. These results are significantly higher than the state-of-the-art reported in Section 2. Also the age results are very high, with values over 70% (e.g., *alvarezcarmona15*, *gonzalesgallardo15*, *kiprov15*, *kocher15*, *sulea15*, and *weren15*). With regard to personality traits, the lowest RMSE was achieved by *kocher15* [30], who used the 200 most frequent terms. The authors obtained also the lowest errors for *Agreeable*, *Conscientious*, and *Openness*, with values pretty close to 10%.

Table 6 shows the results for Italian. Similar to Dutch, the results for this language pertain to gender and personality traits only. As can be seen, results for gender are over 80% in some cases. Very interesting are the results obtained by *gonzalesgallardo15* (0.8611), *grivas15* (0.8333), and *nowson15* [48] (0.8056). The last one applied combinations of n-grams, POS, Named Entities, character flooding, emoticons, and hashtags. With regard to personality traits, *alvarezcarmona15* obtained a significant improvement with an RMSE of about 10%. Moreover, the authors obtained very low results for *Extroversion* (0.0726) and *Agreeable* (0.0527), such as *gonzalesgallardo15* (0.0764 and 0.0745 respectively).

Table 5. Evaluation results in terms of accuracy for age and gender identification (left) and RMSE in personality recognition (right) on **Spanish** texts.

	Age and Gender			**Personality Traits**					
Team	*Joint*	Gender	Age	*RMSE*	E	S	A	C	O
alvarezcarmona15	**0.7727**	**0.9659**	**0.7955**	0.1297	**0.1319**	**0.1631**	0.1113	0.1168	0.1257
ameer15	0.4205	0.6932	0.5341	0.2116	0.2786	0.2806	0.1430	0.1410	0.2145
arroju15	0.4886	0.7500	0.6932	0.1817	0.1980	0.2125	0.1727	0.1785	0.1469
bartoli15	0.3295	0.8523	0.4205	0.1562	0.1701	0.1867	0.1463	0.1320	0.1459
bayot15	0.3636	0.6136	0.5682	0.1773	0.1853	0.2025	0.1593	0.1852	0.1540
cheema15	0.4545	0.8409	0.5682	0.1839	0.1599	0.2479	0.1880	0.1526	0.1712
gimenez15	0.4205	0.6250	0.5682	0.1947	0.2097	0.2440	0.1729	0.1853	0.1617
gonzalesgallardo15	0.7045	0.8977	0.7273	0.1555	0.1406	0.2094	0.1168	0.1709	0.1398
grivas15	0.6818	0.9432	0.6932	0.1876	0.1762	0.1965	0.1557	0.2745	0.1353
kiprov15	0.7273	0.9091	0.7841	0.1495	0.1625	0.1884	0.1249	0.1386	0.1334
kocher15	0.6705	0.8182	0.7386	**0.1235**	0.1373	0.1641	**0.1034**	**0.1017**	**0.1108**
maharjan15	0.5795	0.7955	0.6250	0.2702	0.3008	0.2880	0.2569	0.2357	0.2696
markov15	0.3864	0.6591	0.5114	0.2116	0.1877	0.2644	0.1916	0.2400	0.1742
mccollister15	0.3182	0.6818	0.5000	0.1728	0.1877	0.2098	0.1674	0.1588	0.1403
mezaruiz15	0.4091	0.8295	0.5114	0.1660	0.1729	0.2035	0.1536	0.1473	0.1530
miculicich15	0.6250	0.9205	0.6818	0.1647	0.1856	0.1971	0.1327	0.1402	0.1679
nowson15	0.4886	0.7727	0.6705	0.1598	0.1578	0.2023	0.1358	0.1461	0.1571
poulston15	0.5455	0.8409	0.5909	0.1619	0.1669	0.2285	0.1398	0.1412	0.1329
sulea15	0.6591	0.8750	0.7500	0.1599	0.1703	0.1816	0.1501	0.1559	0.1417
teisseyre15	0.2159	0.5568	0.3636	0.2060	0.1957	0.2446	0.1937	0.2194	0.1768
weren15	0.6932	0.8409	0.7727	0.2034	0.2000	0.2489	0.2003	0.1849	0.1831

Table 6. Evaluation results in terms of accuracy for gender identification (left) and RMSE in personality recognition (right) on **Italian** texts.

Team	Gender	*RMSE*	E	S	A	C	O
alvarezcarmona15	0.7222	**0.1044**	**0.0726**	0.1803	**0.0527**	0.1190	**0.0972**
ameer15	0.5833	0.1723	0.1067	0.2303	0.1462	0.1333	0.2449
arroju15	0.5833	0.1581	0.1480	0.1941	0.1520	0.1345	0.1620
bartoli15	0.5000	0.1405	0.1004	0.1889	0.1386	0.1298	0.1450
bayot15	0.5278	0.1989	0.1928	0.2349	0.1820	0.2173	0.1676
cheema15	0.5278	0.1730	0.1607	0.2205	0.1572	0.1364	0.1900
gimenez15	0.6944	0.1764	0.1394	0.2533	0.1624	0.1247	0.2021
gonzalesgallardo15	**0.8611**	0.1294	0.0764	0.2121	0.0745	0.1269	0.1572
grivas15	0.8333	0.1743	0.1350	0.1930	0.1389	0.2461	0.1586
kocher15	0.7778	0.1259	0.1000	**0.1555**	0.1302	**0.1093**	0.1344
maharjan15	0.6944	0.2122	0.1610	0.2181	0.2118	0.2225	0.2476
mccollister15	0.5556	0.1526	0.1296	0.1993	0.1471	0.1263	0.1610
mezaruiz15	0.5000	0.1636	0.1336	0.1997	0.1463	0.1553	0.1831
miculicich15	0.6389	0.1506	0.1093	0.1650	0.1202	0.1683	0.1900
nowson15	0.8056	0.1515	0.0905	0.2147	0.1237	0.1598	0.1686
poulston15	0.7500	0.1378	0.1279	0.1923	0.1257	0.1187	0.1243
sulea15	0.6389	0.1370	0.1141	0.1913	0.1220	0.1140	0.1438
teisseyre15	0.4167	0.2119	0.1616	0.2646	0.2173	0.1764	0.2398
weren15	0.5833	0.1732	0.1143	0.2593	0.1394	0.1344	0.2186

Finally, Table 7 shows the results for Dutch. Outstanding are the accuracies for gender identification achieved by *alvarezcarmona15* (0.9375), *gonzalesgallardo15* (0.9375), and *grivas15* (0.9688). Regarding personality traits, a noticeable result is the "no error" achieved by *alvarezcarmona15* for *Agreeable*. Altogether, there are several results below 10%: *alvarezcarmona15* in *Extroversion* (0.0750), *Openness* (0.0354) or *Stable* (0.0637), *gonzalesgallardo15* (0.0661) for the last trait. Also note that he lowest RMSE (0.0563, achieved by *alvarezcarmona15*) is more than 3% lower than the one obtained by *gonzalesgallardo15* (0.0890).

Table 7. Evaluation results in terms of accuracy for gender identification (left) and RMSE in personality recognition (right) on **Dutch** texts.

Team	Gender	*RMSE*	E	S	A	C	O
alvarezcarmona15	0.9375	**0.0563**	**0.0750**	**0.0637**	**0.0000**	0.1075	**0.0354**
ameer15	0.5938	0.1418	0.1677	0.1686	0.1436	0.1425	0.0866
arroju15	0.5313	0.1627	0.1573	0.2235	0.1672	0.1553	0.1103
bartoli15	0.7188	0.1156	0.1467	0.1393	0.1261	0.0962	0.0696
bayot15	0.5625	0.1863	0.1705	0.2031	0.1631	0.1978	0.1969
cheema15	0.4688	0.1242	0.1369	0.1768	0.0919	0.1237	0.0919
gimenez15	0.7188	0.1607	0.1829	0.1785	0.1705	0.1392	0.1323
gonzalesgallardo15	0.9375	0.0890	0.0901	0.0661	0.0952	0.1299	0.0637
grivas15	**0.9688**	0.1571	0.1467	0.1711	0.1427	0.2278	0.0973
kocher15	0.8125	0.1186	0.1346	0.1225	0.1311	0.1299	0.0750
maharjan15	0.7813	0.2488	0.2102	0.2821	0.2781	0.2378	0.2358
markov15	0.5313	0.1716	0.1768	0.2411	0.1714	0.1436	0.1250
mccollister15	0.8125	0.1419	0.1499	0.1745	0.1497	0.1442	0.0913
mezaruiz15	0.5000	0.1595	0.1604	0.1928	0.1598	0.1787	0.1055
miculicich15	0.8125	0.1175	0.1199	0.1287	0.1046	0.1358	0.0984
nowson15	0.7813	0.1015	0.1350	0.1315	0.1086	**0.0619**	0.0703
poulston15	0.5000	0.1409	0.1752	0.1511	0.1444	0.1344	0.0993
sulea15	0.8438	0.1164	0.1310	0.1405	0.1114	0.1147	0.0846
teisseyre15	0.5938	0.1853	0.1862	0.2107	0.2187	0.1630	0.1479
weren15	0.6563	0.1491	0.1521	0.1620	0.1928	0.1323	0.1061

6 Conclusion

This paper overviews the results of the 3rd International Author Profiling Task at PAN-2015, hosted at CLEF-2015. Given the languages English, Spanish, Italian, and Dutch, 22 participants had to identify gender and personality traits as well as age classes (English and Spanish only). The participants used content-based features (bag of words, words n-grams, term vectors, TF-IDF n-grams, named entities, dictionary words, slang words, ironic words, sentiment words, emotional words) and style-based features (frequencies, punctuations, POS, verbosity measures and many different statistics besides Twitter specific ones such as mentions, hashtags, and urls).

The highest accuracies in gender identification were achieved in Dutch and Spanish with values over 95%. In comparison to previous years of PAN, the systems achieved significantly higher accuracy values for both age and gender identification. This may suggest that, irrespective the shorter length of individual tweets and their informality, the number of tweets per author is sufficient to profile age and gender with high accuracy.

With regard to personality traits, the lowest errors were obtained for Dutch and Italian, with values below 5% for most traits. The *Stable* trait appears the most difficult one to be predicted.

Regarding the features it is difficult to highlight the most important ones, simply because the high number of different ones used and combined by the participants. This year again the Second Order Representation proposed by *alvarezcarmona15* obtained the best results. However; representations based on n-grams, such as the one proposed by *gonzalesgallardo15* or by *grivas15*, were ranked among the top three in every language.

Acknowledgements The PAN task on author profiling has been organised in the framework of the WIQ-EI IRSES project (Grant No. 269180), within the FP 7 Marie Curie People Framework of the European Commission. The work of the first author was partially funded by Autoritas Consulting SA and by Ministerio de Economía y Competitividad de España under grant ECOPORTUNITY IPT-2012-1220-430000. The second author has received funding from the European Union - Seventh Framework Programme (FP7/2007-2013) under grant agreement 610916 - SENSEI. The work of the third author was done within the framework of the DIANA-APPLICATIONS-Finding Hidden Knowledge in Texts: Applications (TIN2012-38603-C02-01) project and the VLC/CAMPUS Microcluster on Multimodal Interaction in Intelligent Systems. We like to thank Meaning Cloud for sponsoring the award for the winner team.

Bibliography

1. Álvarez-Carmona, M.A., López-Monroy, A.P., Montes-Y-Gómez, M., Villaseñor-Pineda, L., Jair-Escalante, H.: Inaoe's participation at pan'15: Author profiling task—notebook for pan at clef 2015. In: Cappellato et al. [10]
2. Aran, O., Gatica-Perez, D.: Cross-domain personality prediction: from video blogs to small group meetings. In: Proceedings of the 15th ACM on International conference on multimodal interaction. pp. 127–130. ACM (2013)
3. Argamon, S., Koppel, M., Fine, J., Shimoni, A.R.: Gender, genre, and writing style in formal written texts. TEXT 23, 321–346 (2003)
4. Arroju, M., Hassan, A., Farnadi, G.: Age, gender and personality recognition using tweets in a multilingual setting—notebook for pan at clef 2015. In: Cappellato et al. [10]
5. Bartoli, A., De-Lorenzo, A., Laderchi, A., Medvet, E., Tarlao, F.: An author profiling approach based on language-dependent content and stylometric features—notebook for pan at clef 2015. In: Cappellato et al. [10]
6. Batrinca, L.M., Mana, N., Lepri, B., Pianesi, F., Sebe, N.: Please, tell me about yourself: automatic personality assessment using short self-presentations. In: Proceedings of the 13th international conference on multimodal interfaces. pp. 255–262. ACM (2011)

7. Bayot, R., Gonçalves, T., Quaresma, P.: Author profiling of twitter users—notebook for pan at clef 2015. In: Cappellato et al. [10]
8. Burger, J.D., Henderson, J., Kim, G., Zarrella, G.: Discriminating gender on twitter. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. pp. 1301–1309. EMNLP '11, Association for Computational Linguistics, Stroudsburg, PA, USA (2011)
9. Cappellato, L., Ferro, N., Halvey, M., Kraaij, W. (eds.): CLEF 2015 Labs and Workshops, Notebook Papers. CEUR-WS.org vol. 1180 (2014)
10. Cappellato, L., Ferro, N., Jones, G., San-Juan, E. (eds.): CLEF 2015 Labs and Workshops, Notebook Papers. CEUR-WS.org vol. 1391 (2015)
11. Celli, F., Bruni, E., Lepri, B.: Automatic personality and interaction style recognition from facebook profile pictures. In: Proceedings of the ACM International Conference on Multimedia. pp. 1101–1104. ACM (2014)
12. Celli, F., Lepri, B., Biel, J.I., Gatica-Perez, D., Riccardi, G., Pianesi, F.: The workshop on computational personality recognition 2014. In: Proceedings of the ACM International Conference on Multimedia. pp. 1245–1246. ACM (2014)
13. Celli, F., Polonio, L.: Relationships between personality and interactions in facebook. In: Social Networking: Recent Trends, Emerging Issues and Future Outlook, pp. 41–54. Nova Science Publishers, Inc (2013)
14. Chittaranjan, G., Blom, J., Gatica-Perez, D.: Mining large-scale smartphone data for personality studies. Personal and Ubiquitous Computing 17(3), 433–450 (2013)
15. Costa, P.T., McCrae, R.R.: The revised neo personality inventory (neo-pi-r). The SAGE handbook of personality theory and assessment 2, 179–198 (2008)
16. Fornaciari, T., Celli, F., Poesio, M.: The effect of personality type on deceptive communication style. In: Intelligence and Security Informatics Conference (EISIC), 2013 European. pp. 1–6. IEEE (2013)
17. Forner, P., Navigli, R., Tufis, D. (eds.): CLEF 2013 Evaluation Labs and Workshop – Working Notes Papers. CEUR-WS.org vol. 1179 (2013)
18. Gimenez, M., Irazú-Hernández, Plá, F.: Segmenting target audiences: Authomatic author profiling using tweets—notebook for pan at clef 2015. In: Cappellato et al. [10]
19. Golbeck, J., Robles, C., Turner, K.: Predicting personality with social media. In: CHI'11 Extended Abstracts on Human Factors in Computing Systems. pp. 253–262. ACM (2011)
20. Gollub, T., Potthast, M., Beyer, A., Busse, M., Rangel, F., Rosso, P., Stamatatos, E., Stein, B.: Recent Trends in Digital Text Forensics and its Evaluation. In: Forner, P., Müller, H., Paredes, R., Rosso, P., Stein, B. (eds.) Information Access Evaluation meets Multilinguality, Multimodality, and Visualization. 4th International Conference of the CLEF Initiative (CLEF 13). pp. 282–302. Springer, Berlin Heidelberg New York (Sep 2013)
21. Gollub, T., Stein, B., Burrows, S.: Ousting Ivory Tower Research: Towards a Web Framework for Providing Experiments as a Service. In: Hersh, B., Callan, J., Maarek, Y., Sanderson, M. (eds.) 35th International ACM Conference on Research and Development in Information Retrieval (SIGIR 12). pp. 1125–1126. ACM (Aug 2012)

22. Gollub, T., Stein, B., Burrows, S., Hoppe, D.: TIRA: Configuring, Executing, and Disseminating Information Retrieval Experiments. In: Tjoa, A.M., Liddle, S., Schewe, K.D., Zhou, X. (eds.) 9th International Workshop on Text-based Information Retrieval (TIR 12) at DEXA. pp. 151–155. IEEE, Los Alamitos, California (Sep 2012)
23. González-Gallardo, C.E., Montes, A., Sierra, G., Núñez, A., Adolfo, S., Ek, J.: Tweets classification using corpus dependent tags, character and pos n-grams—notebook for pan at clef 2015. In: Cappellato et al. [10]
24. Grivas, A., Krithara, A., Giannakopoulos, G.: Author profiling using stylometric and structura feature groupings—notebook for pan at clef 2015. In: Cappellato et al. [10]
25. Guyon, I., Wston, J., Barnhill, S., Vapnik, V.: Gene selection for cancer classification using support vector machines. In: Machine Learning 46 (1-3). pp. 389–422 (2002)
26. Holmes, J., Meyerhoff, M.: The Handbook of Language and Gender. Blackwell Handbooks in Linguistics, Wiley (2003)
27. Houvardas, J., Stamatatos, E.: N-gram feature selection for authorship identification. In: Artificial Intelligence: Methodology, Systems, and Applications. Springer. pp. 77–86 (2006)
28. Kalimeri, K., Lepri, B., Pianesi, F.: Going beyond traits: Multimodal classification of personality states in the wild. In: Proceedings of the 15th ACM on International Conference on Multimodal Interaction. pp. 27–34. ICMI '13, ACM, New York, NY, USA (2013)
29. Kiprov, Y., Hardalov, M., Nakov, P., Koychev, I.: Su@pan'2015: Experiments in author profiling—notebook for pan at clef 2015. In: Cappellato et al. [10]
30. Kocher, M.: Unine at clef 2015: Author profiling—notebook for pan at clef 2015. In: Cappellato et al. [10]
31. Koppel, M., Argamon, S., Shimoni, A.R.: Automatically categorizing written texts by author gender. literary and linguistic computing 17(4) (2002)
32. Kosinski, M., Bachrach, Y., Kohli, P., Stillwell, D., Graepel, T.: Manifestations of user personality in website choice and behaviour on online social networks. Machine Learning pp. 1–24 (2013)
33. Lepri, B., Subramanian, R., Kalimeri, K., Staiano, J., Pianesi, F., Sebe, N.: Connecting meeting behavior with extraversion: A systematic study. Affective Computing, IEEE Transactions on 3(4), 443–455 (2012)
34. López-Monroy, A.P., y Gómez, M.M., Jair-Escalante, H., nor Pineda, L.V.: Using Intra-Profile Information for Author Profiling—Notebook for PAN at CLEF 2014. In: Cappellato et al. [9]
35. Lopez-Monroy, A.P., Montes-Y-Gomez, M., Escalante, H.J., Villasenor-Pineda, L., Villatoro-Tello, E.: INAOE's Participation at PAN'13: Author Profiling task—Notebook for PAN at CLEF 2013. In: Forner et al. [17]
36. Maharjan, S., Shrestha, P., Solorio, T., Hasan, R.: A straightforward author profiling approach in mapreduce. In: Advances in Artificial Intelligence. Iberamia. pp. 95–107 (2014)
37. Maharjan, S., Solorio, T.: Using wide range of features for author profiling—notebook for pan at clef 2015. In: Cappellato et al. [10]

38. Mairesse, F., Walker, M.A., Mehl, M.R., Moore, R.K.: Using linguistic cues for the automatic recognition of personality in conversation and text. Journal of Artificial Intelligence Research 30(1), 457–500 (2007)
39. McCollister, C., Luo, B., Huang, S.: Building topic models to predict author attributes from twitter messages—notebook for pan at clef 2015. In: Cappellato et al. [10]
40. Meina, M., Brodzinska, K., Celmer, B., Czokow, M., Patera, M., Pezacki, J., Wilk, M.: Ensemble-based Classification for Author Profiling Using Various Features—Notebook for PAN at CLEF 2013. In: Forner et al. [17]
41. Miculicich, L.: Statistical learning methods for profiling analysis—notebook for pan at clef 2015. In: Cappellato et al. [10]
42. Mohammad, S.: #emotional tweets. In: *SEM 2012: The First Joint Conference on Lexical and Computational Semantics. pp. 246–255 (2012)
43. Mohammadi, G., Vinciarelli, A.: Automatic personality perception: Prediction of trait attribution based on prosodic features. Affective Computing, IEEE Transactions on 3(3), 273–284 (2012)
44. de Montjoye, Y.A., Quoidbach, J., Robic, F., Pentland, A.S.: Predicting personality using novel mobile phone-based metrics. In: Social Computing, Behavioral-Cultural Modeling and Prediction, pp. 48–55. Springer (2013)
45. Najib, F., Cheema, A., Muhammad, R., Nawab, A.: Author's traits prediction on twitter data using content based approach—notebook for pan at clef 2015. In: Cappellato et al. [10]
46. Nguyen, D., Gravel, R., Trieschnigg, D., Meder, T.: "how old do you think i am?"; a study of language and age in twitter. Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media (2013)
47. Nguyen, D., Smith, N.A., Rosé, C.P.: Author age prediction from text using linear regression. In: Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities. pp. 115–123. LaTeCH '11, Association for Computational Linguistics, Stroudsburg, PA, USA (2011)
48. Nowson, S., Perez, J., Brun, C., Mirkin, S., Roux, C.: Xrce personal language analytics engine for multilingual author profiling—notebook for pan at clef 2015. In: Cappellato et al. [10]
49. Oberlander, J., Nowson, S.: Whose thumb is it anyway?: classifying author personality from weblog text. In: Proceedings of the COLING/ACL on Main conference poster sessions. pp. 627–634. Association for Computational Linguistics (2006)
50. Olguın, D.O., Gloor, P.A., Pentland, A.S.: Capturing individual and group behavior with wearable sensors. In: Proceedings of the 2009 aaai spring symposium on human behavior modeling, SSS. vol. 9 (2009)
51. Palomino-Garibay, A., Camacho-González, A.T., Fierro-Villaneda, R.A., Hernández-Farias, I., Buscaldi, D., Meza-Ruiz, I.V.: A random forest approach for authorship profiling. In: Cappellato et al. [10]
52. Peersman, C., Daelemans, W., Vaerenbergh, L.V.: Predicting age and gender in online social networks. In: Proceedings of the 3rd international workshop on Search and mining user-generated contents. pp. 37–44. SMUC '11, ACM, New York, NY, USA (2011)

53. Pennebaker, J.W.: The Secret Life of Pronouns: What Our Words Say About Us. Bloomsbury USA (2013)
54. Pennebaker, J.W., Mehl, M.R., Niederhoffer, K.G.: Psychological aspects of natural language use: Our words, our selves. Annual review of psychology 54(1), 547–577 (2003)
55. Pervaz, I., Ameer, I., Sittar, A., Muhammad, R., Nawab, A.: Identification of author personality traits using stylistic features—notebook for pan at clef 2015. In: Cappellato et al. [10]
56. Pianesi, F., Mana, N., Cappelletti, A., Lepri, B., Zancanaro, M.: Multimodal recognition of personality traits in social interactions. In: Proceedings of the 10th international conference on Multimodal interfaces. pp. 53–60. ACM (2008)
57. Posadas-Durán, J.P., Markov, I., Gómez-Adorno, H., Sidorov, G., Batyrshin, I., Gelbukh, A., Pichardo-Lagunas, O.: Syntactic n-grams as features for the author profiling task—notebook for pan at clef 2015. In: Cappellato et al. [10]
58. Poulston, A., Stevenson, M., Bontcheva, K.: Topic models and n-gram language models for author profiling—notebook for pan at clef 2015. In: Cappellato et al. [10]
59. Przybyla, P., Teisseyre, P.: What do your look-alikes say about you? explotting strong and wak similarities for author profiling—notebook for pan at clef 2015. In: Cappellato et al. [10]
60. Quercia, D., Lambiotte, R., Stillwell, D., Kosinski, M., Crowcroft, J.: The personality of popular facebook users. In: Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work. pp. 955–964. ACM (2012)
61. Rammstedt, B., John, O.: Measuring personality in one minute or less: A 10 item short version of the big five inventory in english and german. In: Journal of Research in Personality. pp. 203–212 (2007)
62. Rangel, F., Rosso, P.: On the impact of emotions on author profiling. In: Information Processing & Management, Special Issue on Emotion and Sentiment in Social and Expressive Media (In Press) (2015)
63. Rangel, F., Rosso, P., Chugur, I., Potthast, M., Trenkmann, M., Stein, B., Verhoeven, B., Daelemans, W.: Overview of the 2nd author profiling task at pan 2014. In: In: Cappellato L., Ferro N., Halvey M., Kraaij W. (Eds.) CLEF 2014 Labs and Workshops, Notebook Papers. CEUR-WS.org, vol. 1180 (2014)
64. Rangel, F., Rosso, P., Koppel, M., Stamatatos, E., Inches, G.: Overview of the Author Profiling Task at PAN 2013—Notebook for PAN at CLEF 2013. In: Forner et al. [17]
65. Rizwan-Iqbal, H., Adnan-Ashraf, M., Adeel-Nawab, R.M.: Predicting an author's demographics from text using topic modeling approach—notebook for pan at clef 2015. In: Cappellato et al. [10]
66. Roshchina, A., Cardiff, J., Rosso, P.: A comparative evaluation of personality estimation algorithms for the twin recommender system. In: Proceedings of the 3rd international workshop on Search and mining user-generated contents. pp. 11–18. ACM (2011)
67. Schartz, A., Eichstaedt, J., Kern, M., Dziurzynski, L., Ramones, S., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Scligman, M., Ungar, L.: Personality, gender and age in the language of social media: The open-vocabulary approach. In: In: PLOS ONE. Public Library of Science. p. e73791 (2013)

68. Schler, J., Koppel, M., Argamon, S., Pennebaker, J.W.: Effects of age and gender on blogging. In: AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs. pp. 199–205. AAAI (2006)
69. Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Dziurzynski, L., Ramones, S.M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M.E., et al.: Personality, gender, and age in the language of social media: The open-vocabulary approach. PloS one 8(9), 773–791 (2013)
70. Staiano, J., Lepri, B., Aharony, N., Pianesi, F., Sebe, N., Pentland, A.: Friends don't lie: inferring personality traits from social network structure. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing. pp. 321–330. ACM (2012)
71. Sulea, O.M., Dichiu, D.: Automatic profiling of twitter users based on their tweets—notebook for pan at clef 2015. In: Cappellato et al. [10]
72. Sushant, S.A., Argamon, S., Dhawle, S., Pennebaker, J.W.: Lexical predictors of personality type. In: In Proceedings of the Joint Annual Meeting of the Interface and the Classification Society of North America (2005)
73. Vinciarelli, A., Mohammadi, G.: A survey of personality computing. IEEE Transactions on Affective Computing 5(3), 1–1 (2014)
74. Weren, E.: Information retrieval features for personality traits—notebook for pan at clef 2015. In: Cappellato et al. [10]
75. Weren, E., Kauer, A., Mizusaki, L., Moreira, V., de Oliveira, P., Wives, L.: Examining multiple features for author profiling. In: Journal of Information and Data Management. pp. 266–279 (2014)
76. Youyou, W., Kosinski, M., Stillwell, D.: Computer-based personality judgments are more accurate than those made by humans. PNAS pp. 1–5 (2015)
77. Zhang, C., Zhang, P.: Predicting gender from blog posts. Tech. rep., Technical Report. University of Massachusetts Amherst, USA (2010)