

# Domain Independent Authorship Attribution without Domain Adaptation

**Rohith K Menon**

Department of Computer Science  
Stony Brook University  
rkmenon@cs.stonybrook.edu

**Yejin Choi**

Department of Computer Science  
Stony Brook University  
ychoi@cs.stonybrook.edu

## Abstract

Automatic authorship attribution, by its nature, is much more advantageous if it is domain (i.e., topic and/or genre) independent. That is, many real world problems that require authorship attribution may not have in-domain training data readily available. However, most previous work based on machine learning techniques focused only on in-domain text for authorship attribution. In this paper, we present comprehensive evaluation of various stylometric techniques for cross-domain authorship attribution. From the experiments based on the Project Gutenberg book archive, we discover that extremely simple techniques based on stopwords are surprisingly robust against domain change, essentially ridding the need for domain adaptation when supplied with a large amount of data.

## 1 Introduction

Many real world problems that require authorship attribution, such as forensics (e.g., Luyckx and Daelemans (2008)) or authorship dispute for old literature (e.g., Mosteller and Wallace (1984)) may not have in-domain training data readily available. However, most previous work to date has focused on authorship attribution only for in-domain text (e.g., Stamatatos et al. (1999), Luyckx and Daelemans (2008), Raghavan et al. (2010)). On limited occasions researchers include heterogeneous (cross-domain) dataset in their experiments, but they only report the performance on heterogeneous dataset is much lower than that of homogeneous dataset, rather than directly tackling the problem of cross-domain or domain independent authorship attribution (e.g., Peng et al. (2003)).

The lack of research for cross-domain scenarios is perhaps only reasonable, given that it is understood in the community that the prediction power

of machine learning techniques does not transfer well over different domains (e.g., Blitzer et al. (2008)). However, the seminal work of Blitzer et al. (2006) has shown that it is possible to mitigate the problem by examining distributional difference of features across different domains, and derive features that are robust against domain switch. Therefore, one could expect that applying domain adaptation techniques to authorship attribution can also help with cross-domain authorship attribution.

Before hastening into domain adaptation for authorship attribution, we take a slightly different push to the problem: we first examine whether there exist domain-independent features that rarely change across different domains. If this is the case, and if such features are sufficiently informative, then domain adaptation might not be required at all to achieve high performance in domain-independent authorship attribution. Therefore, we conduct a comprehensive empirical evaluation using various stylistic features that are likely to be common across different topics and genre.

From the experiments based on the Project Gutenberg book archive, we indeed discover stylistic features that are common across different domains. Against our expectations, some of such features, stop-words in particular, are extremely informative, essentially ridding of the need for domain adaptation, if supplied with a large amount of data. Due to its simplicity, techniques based on stop-words scale particularly well over a large amount of data, in comparison to more computationally heavy techniques that require parsing (e.g., Raghavan et al. (2010)).

## 2 Domain Independent Cues for Author Identification

The study of authorship attribution requires careful preparation of dataset, in order not to draw

overly optimistic conclusions. For instance, if the dataset consists of text where each author writes about a distinctive and exclusive topic, the task of author attribution reduces to topic categorization, a much easier task in general (e.g., (Mikros and Argiri, 2007)). Such statistical models that rely on topics will not generalize well over text in previously unseen topics or genre. Random collection of data is not the solution to this concern, as many authors are biased toward certain topics and genre. In order to avoid such pitfall of inadvertently benefiting from topic bias, we propose two different ways of data preparation: First approach is to ensure that multiple number of authors are included per topic and genre, so that it is hard to predict the author purely based on topical words. Second approach is to ensure that multiple domains (i.e., topics and/or genre) are included per author, and that test dataset includes domains that are previously unseen in the training data.

Next we discuss stylistic features that are likely to be common across different domains. In this study, we compare the following set of features: (1) n-gram sequences as a baseline, (2) part of speech sequences that capture shallow syntactic patterns, (3) modified  $tf - idf$  for n-gram that captures repeated phrases, (4) mood words that capture author’s unique emotional traits, and (5) stop word frequencies that capture author’s writing habit with common words. Each of these features is elaborated below.

## 2.1 N-gram Sequences as a Topic Dependent Baseline

We conjecture that N-gram sequences are not robust against domain changes, as N-grams are powerful features for topic categorization (e.g., (Türkoğlu et al., 2007)). We therefore set N-gram based features as baseline to quantify how much domain change affects the performance. Normalized frequency of the most frequent 100 stemmed (Porter, 1997) 3-grams<sup>1</sup> are encoded as features.

## 2.2 3-gram Part-of-Speech Sequences to Capture Favorite Sentence Structure

To capture the syntactic patterns unique to authors, we use 3-gram sequence of part-of-speech (POS) tags. To be robust across domain change, we use

<sup>1</sup>For all ngram based features, 3-gram (N=3) was chosen because increasing N increased sparseness and decreasing N failed to capture common phrases.

only the most frequent 100 3-grams of part-of-speech tags as features. To encode a feature from each such 3-gram POS sequence, we use the frequency of each POS sequence normalized by the number of POS grams in the document. We expect these shallow syntactic patterns will help characterize the favorite sentence structure used by the authors. We make use of Stanford parser (Klein and Manning, 2003) to tag the part-of-speech tags for the given document.

## 2.3 Modified $tf - idf$ for 3-gram Sequences

$Tf - idf$  provides a score to a term indicating how informative each term is, by multiplying the frequency of the term within the document (term frequency) by the rarity of the term across corpus (inverse document frequency).  $tf - idf$  is known to be highly effective for text categorization. In this work, we experiment with modified  $tf - idf$  in order to accommodate the nature of author attribution more directly. We propose two such variants:

### ***tf-iAf - Term-Frequency Inverse-Author-Frequency***

In this variant, we take inverse-*author*-frequency instead of inverse-*document*-frequency, as the terms that occur across many authors are not as informative as the terms unique to a given author. For training data, we compute  $tf-iAf$  based on known authors of each document, however in test data, we do not have access to the authors of each document. Therefore, we set  $tf-iAf$  of the test data as  $tf$  of the test data weighted by  $iAf$  of the training data. We generate these features for top 500 3-gram sequences ordered by  $tf-iAf$  scores from each author. We compute different  $tf-iAf$  values for different authors. The exact formula we use for a given author  $i$  is given below:

$$Tfiaf_i = \sum_{j=1}^{K_i} \frac{f_{ij}}{N_{ij}} * iaf_i^2$$

where  $f_{ij}$  is the frequency of a 3-gram for  $author_i$  in document  $D_{ij}$ ,  $D_{ij}$  is the  $j$ th document by  $author_i$ ,  $N_{ij}$  is the total number of 3-grams in document  $D_{ij}$ , and  $K_i$  is the number of documents written by  $author_i$ . We take the second power of inverse-author-frequency, as the number of authors is much smaller than the number of documents in a corpus.

### ***tf-iAf-tpf* – Term-Frequency Inverse-Author-Frequency Topic-Frequency**

In this variant, we augment the previous approach with *topic*-frequency, which is the number of different topics a given term appears for a given author. We generate these features for top 500 3-gram sequences ordered by *tf-iAf-tpf* scores from each author. Again, we compute different *tf-iAf-tpf* values for different authors. The exact formula for a given author  $i$  is given below:

$$TfiafTpf_i = Tfiaf_i * tpf_i^2$$

where we take the second power to the topic frequency, as the number of distinctive topics is small in general.

## **2.4 Mood Words to Capture Emotional Traits**

We conjecture that mood words<sup>2</sup> will reveal unique emotional traits of each author. In particular, either the use of certain types of mood words, or the lack of it, will reveal common mood or tone in documents that is orthogonal to the topics or genre. To encode features based on mood words, we include the normalized frequency of each mood word in a given document in the feature vector. Normalization is done by dividing frequency by total number of words in the document. We consider in total a list of 859 mood words.

## **2.5 Stop-words to Capture Writing Habit**

Many researchers reported that the usage patterns of stop-words are a very strong indication of writing style (Arun et al. (2009), Garca and Martn (2007)). Based on 659 stop words obtained, we encode features as the frequency of each stop-word normalized by total number of words in the document<sup>3</sup>. These normalized frequencies indicate two important characteristics of stop-word usage by authors:

- (1) Relative usage of function words by authors.
- (2) Fraction of function words in document.

## **3 Dataset with Varying Degree of Domain Change**

In order to investigate the topic influence on authorship attribution, we need a dataset that consists

of articles written by prolific authors who wrote on a variety of topics. Furthermore, it would be ideal if the dataset already includes topic categorization, so that we do not need to manually categorize each article into different topics and genre.

Fortunately, there is such a dataset available online: we use the project Gutenberg book archive (<http://www.gutenberg.org>) that contains an extensive collection of books. In order to remove topic bias in authors, we rely on the catalog of project Gutenberg. Categories of project Gutenberg correspond to the mixture of topics and genre.

There are two types of categories defined in project Gutenberg: the first is LCSH (Library of Congress Subject Headings)<sup>4</sup> and the second is LCC (Library of Congress Classification).<sup>5</sup> Examples of LCSH and LCC categories are shown in Table 1 and Table 2 respectively. As can be seen in Table 1, the categories of LCSH are more fine-grained, and some of the categories are overlapping eg: “*history*” and “*history and criticism*”. In contrast, the categories of LCC are more coarse-grained so that they are more distinctive from each other.

In the next section, we present following four experiments in the order of increasing difficulty. We use the term topics, genre, and domains interchangeably in what follows, as LCC & LCSH categories are mixed as well.

- (1) **Balanced topic:** Topics in the test data are guaranteed to appear in the training data.
- (2) **Semi-disjoint topic using LCSH:** Topics in the test data differ from topics in the training according to LCSH.
- (3) **Semi-disjoint topic using LCC:** Topics in the test data differ from topics in the training according to LCC.
- (4) **Perfectly-disjoint topic using LCC:** Topics in the test data differ from topics in the training according to LCC, and documents with unknown categories are discarded to create perfectly disjoint training and test data, while in (2) and (3) documents with unknown categories are added to maintain large dataset.

<sup>2</sup>The list of mood words is obtained from <http://moods85.wordpress.com/mood-list/>

<sup>3</sup>The list of stopwords is obtained from <http://www.ranks.nl/resources/stopwords.html>

<sup>4</sup><http://www.loc.gov/aba/cataloging/subject/weeklylists/>

<sup>5</sup><http://www.loc.gov/catdir/cpsol/lcco/>

American drama, Eugenics, American poetry, Fairy tales, Architecture, Family, Art, Farm life, Authors, Fiction, Ballads, Fishing, Balloons, France, Children, Harbors, Civil War, History, Conduct of life, History and criticism, Correspondence, History – Revolution, Country life, Cycling, Description and travel, ...

Table 1: Examples of LCSH Categories.

Music And Books On Music, Philosophy, Psychology, Fine Arts, Religion, Auxiliary Sciences Of History, Language And Literature, World History (Non Americas), Science, History Of The Americas, Medicine, Geography, Anthropology, Agriculture, Recreation, Social Sciences, Technology, Political Science, ...

Table 2: Examples of LCC Categories.

Author	Total	LCC	LCSH
Andrew Lang	63	(36, 8)	(16, 12)
Charles Kingsley	45	(10, 6)	(2, 2)
Charlotte Mary	59	(27, 5)	(11, 9)
G K Chesterton	37	(22, 7)	(7, 6)
H G Wells	43	(38, 7)	(12, 10)
Jacob Abbott	48	(33, 9)	(15, 14)
John Morley	27	(8, 5)	(6, 6)
John Ruskin	38	(16, 8)	(8, 7)
R M Ballantyne	97	(85, 9)	(5, 5)
Robert Louis	80	(28, 2)	(19, 6)
Thomas Carlyle	35	(6, 5)	(1, 1)
Thomas Henry	41	(12, 4)	(4, 3)
William Dean	95	(38, 6)	(25, 19)
William Henry	113	(24, 4)	(2, 2)

Table 3: Author statistics. Numbers in parentheses ( $x, y$ ) under LCC and LCSH columns indicate the number of books categorized ( $x$ ) and the number of unique categories the author has written in ( $y$ ).

## 4 Experimental Results

We present four experiments in the order of increasing difficulty. In all experiments, we use the SVM classifier with sequential minimal optimization (SMO) implementation available in the Weka package (Hall et al., 2009). We used polynomial kernel with regularization parameter  $C = 1$ .

### 4.1 Balanced Topic

**Configuration** We identify a set of 14 authors who had written at least 25 books and also had written books in at least 6 categories. This amounts to 844 books in total for all authors. Table 3 tabulates the author statistics.

In our first experiment, we randomly split the 844 books into 744 training data and 100 testing data with 14 authors. This setting is simpler than true topic disjoint scenario where there is no intersection between topics in training and testing sets. Nevertheless, this setting is not an easy one, as we only consider authors who have written for more than 6 topics, which makes it harder to benefit from topic bias in authors. Note that a random guess will give an accuracy of  $\frac{1}{14}$  only.

**Result** Table 4 tabulates the accuracy, precision, recall and f-score obtained for various features described in Section 2. Note that f-scores (including precision and recall) are first computed for each author, then we take the macro average over different authors. We perform 8-way cross validation for this setup. The first row — N-GRAM — is the baseline. It is interesting that n-gram-based features suffer in this experimental setting already, even though we do not deliberately change the topics across training and test data. All other features

Features	Acc	Prec	Rec	F1
NGram	61.22	64.75	59.51	58.02
Tflaf	90.82	94.69	91.54	92.10
TflafTpf	84.69	86.02	85.61	84.96
POSGram	91.84	93.19	91.22	91.51
MoodWord	95.92	94.99	96.28	95.22
StopWord	<b>97.96</b>	<b>99.21</b>	<b>97.92</b>	<b>98.45</b>
All	93.88	95.30	94.68	94.41

Table 4: Balanced Topic (Experiment-1)

demonstrate strong performance, mostly achieving F-score and accuracy well above 90%, with the exception of TflafTpf.

Stop-word based features achieve the highest performance with 98.45% in F-score and 97.96% in accuracy. This echoes previously reported studies (e.g., Arun et al. (2009)) that indicate that stop words can reveal author’s unique writing styles and habits. We are nonetheless surprised to see the performance of stopword based features is higher than that of more sophisticated approaches such as Tflaf or TflafTpf.

It is unexpected to see that tfiaf-tpf performs worse than tfiaf or POS-grams. We conjecture the cause can be attributed to the fact that we calculate tfiaf-tpf only from the set of books which are categorized by LCC. We calculate tfiaf-tpf only from LCC categorized books because only these categories at the root level are truly disjoint. Because

we select tfiaf-tpf ngrams only from the subset of the books in training, it is possible that we could have missed some ngrams which would otherwise have high tfiaf-tpf scores.

High performance for mood words, reaching 95.22% in F-score and 95.92% in accuracy confirms our hypothesis that it can reveal author’s unique emotional traits that are orthogonal to particular topics.

**Note on the Baseline** Because the baseline scores are very low, we also experimented with other variants with baselines not included in the table for brevity. First, we tested with increased number of n-grams. That is, instead of using top 100 3-grams per document, we experiment with top 500 3-grams per document. This did not change the performance much however. We also tried to incorporate all 3-grams, but we could not fit such features based on all 3-grams into memory, as our dataset consists of many books in their entirety. We conclude the discussion on the first experiment by highlighting two important observations:

- First, POS 3-gram features are also based on top 100 POS 3-grams per document, and *these unlexicalized features perform extremely well with 91.51% f-score and 91.84% accuracy, using the identical number of features as the baseline.*
- Second, all features presented here are highly *efficient* and *scalable*.

## 4.2 Semi-Disjoint Topic using LCSH

**Configuration** In the second experiment, we use categories from LCSH. As shown in Table 1, these categories were not completely disjoint. As a result, we split training and test data with manual inspection on the LCSH categories to ensure training and test data are as disjoint as possible. In this experiment, we focus on 6 authors out of 14 authors considered in the previous dataset in order to make it easier to split training and test data based on disjoint topics. In particular, we place books in fiction, essays and history categories in the training set, and the rest in the test set. This results in 202 books for training and 72 books for testing.

Despite our effort, this split is not perfect: first, it might still allow topics with very subtle differences to show up in both training and test data. Second, the training set includes books that are not categorized by LCSH categories. As a re-

Features	Acc	Prec	Rec	F1
NGram	52.78	57.69	53.61	52.66
TfIaf	87.50	89.73	86.15	84.53
TfIafTpf	81.94	82.29	80.22	79.47
POSGram	86.11	88.89	84.81	85.57
MoodWord	87.50	88.28	84.90	85.77
StopWord	<b>98.61</b>	<b>98.81</b>	<b>98.72</b>	<b>98.72</b>
All	93.06	94.23	92.44	92.47

Table 5: Semi-Disjoint Topics using LCSH (Experiment-2)

sult, these books with unknown categories might accidentally contain books whose topics overlap with the topics included in the test data. Nevertheless, author attribution becomes a much harder task than before, because a significant portion of training and test data consists of disjoint topics.

**Result** Table 5 tabulates the results. As expected, the overall performance drops for almost all approaches. The only exceptional case is stop word based features, the top performer in the previous experiment. It is astonishing that the performance of stop word based features in fact does not drop at all, achieving 98.72% in f-score and 98.61% in accuracy. As before, the mixture of all features actually decrease the performance. Overall the performance of most approaches look strong however, as most achieve scores well above 80% in f-score and accuracy. Baseline performs very poorly again, as n-grams are more sensitive to topic changes than other features.

## 4.3 Semi-Disjoint Topic using LCC

**Configuration** For the third experiment, we use categories from LCC instead of LCSH. As described earlier, top categories of LCC are more disjoint than those of LCSH. We choose 5 authors who have written in "Language and literature" in addition to other categories. We then create a training set with books in categories that are not "Language and Literature". We also include books with unknown categories into the training dataset to maintain a reasonably large dataset. The test set consists of books in a single topic "Language and Literature". This split results in 146 books for training, and 112 books for testing.

**Result** Table 6 tabulates the result. Again, the f-score (including precision and recall) are first computed per-author, then we take the macro aver-

Features	Acc	Prec	Rec	F1
NGram	70.54	70.95	64.88	65.84
TfIaf	93.75	95.66	89.76	91.37
TfIafTpf	88.39	91.89	82.18	83.40
POSGram	93.75	94.80	89.23	90.14
MoodWord	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>
StopWord	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>
All	98.21	98.67	96.67	97.51

Table 6: Semi-Disjoint Topics using LCC (Experiment-3)

age over all authors. Surprisingly, the performance of all approaches increased. We conjecture the reason to be overlap of unknown categories with categories in the test dataset.

Stop word and mood based features achieve 100% prediction accuracy in this setting. However, we should like to point out that this extremely high performance of simple features are attainable only when supplied with sufficiently large amount of data. See Section 4.5 for discussions related to the performance change with reduced data size.

#### 4.4 Perfectly-Disjoint Topic using LCC

**Configuration** Finally, we experiment on a set of data which were truly topic independent, and we try to learn the author cues from one topic and use it to predict the authors of books written in different topics. In this experiment, the training set consists of books in a single topic "Language and Literature", which used to be the test dataset in the previous experiment. For test, we take the training dataset from the previous experiment and remove those books with unknown categories to enforce fully disjoint topics between training and testing. This split results in 112 documents in the training data and 37 documents in the test data.

**Result** Table 7 tabulates the result. Note that this experiment is indeed the harder than the previous experiment, as the performance of the most approaches dropped significantly. Here we find that the performance of tfIaf-tpf is very strong achieving 95.33% in f-score and 94.59% in accuracy. Note that in all of previous experiments, tfIaf-tpf performed considerably worse than tfIaf. This is because this experiment is the only experiment that discards all books with unknown categories, which makes it possible for tfIaf-tpf to exploit the topic information more accurately. In

Features	Acc	Prec	Rec	F1
NGram	56.76	55.33	55.50	53.07
TfIaf	86.49	89.00	89.39	87.35
TfIafTpf	94.59	95.00	96.36	95.33
POSGram	64.86	69.57	71.17	69.33
MoodWord	81.08	83.83	83.12	81.84
StopWord	<b>97.30</b>	<b>97.50</b>	97.14	97.13
All	<b>97.30</b>	<b>97.50</b>	<b>98.18</b>	<b>97.71</b>

Table 7: Perfectly-Disjoint Topics using LCC (Experiment-4)

fact, the performance of tfIaf-tpf is now almost as good as that of stop word based features, our all time top performer that achieves 97.13% in f-score and 97.30% in accuracy in this experiment. Mood words and pos-grams, previously high performing approaches do not appear to be very robust with drastic domain changes.

#### 4.5 Perfectly-Disjoint Topic using LCC with Reduced Data

In this section, we briefly report how the performance of all approaches changes when we reduce the size of the data. For brevity, we report this only with respect to the last experiment. Table 8 shows the results, when we reduce the size of data down to 10% and 50% respectively, by taking the first  $x\%$  of each book in the training and test data. In comparison to Table 7, overall performance drops with reduced data. From these results, we conclude that (1) when faced with data reduction, the relative performance of stop word based features stands out even more, and that (2) high performance of simple features are attainable when supplied with sufficiently large amount of data.

## 5 Related Work

Stamatatos (2009) provides an excellent survey of the field. One of the prominent approaches in authorship attribution is the use of style markers (Stamatatos et al., 1999). Our approaches make use of such style markers implicitly and more systematically.

The work of Peng et al. (2003) by using character level n-grams achieve state-of-the-art performance (90%) on homogeneous (in-domain) but drops significantly (74%) on heterogeneous (cross-domain) data in accuracy. In contrast, we present approaches that perform extremely well even on heterogeneous data.

Features	Acc <sub>10</sub>	F1 <sub>10</sub>	Acc <sub>50</sub>	F1 <sub>50</sub>
NGram	37.84	39.40	48.65	46.78
TfIaf	32.43	30.57	72.97	75.43
TfIafTpf	32.43	33.11	62.16	62.87
POSGram	24.32	31.16	62.16	64.23
MoodWord	40.54	36.77	70.27	67.10
StopWord	<b>64.86</b>	<b>65.38</b>	<b>91.89</b>	<b>92.12</b>
All	37.84	39.24	75.68	77.01

Table 8: Perfectly-Disjoint Topics using LCC (Reduced to 10% and 50% of the original data)

Another interesting technique that is explored for authorship attribution is the use of PCFG in the work of Raghavan et al. (2010). They show that PCFG models are effective in authorship attribution, although their experiments were conducted only on homogeneous datasets. The approaches studied in this paper are much simpler and highly scalable, while extremely effective.

## 6 Conclusion

We have presented a set of features for authorship attribution in a domain independent setting. We have demonstrated that the features we calculate are effective in predicting authorship while being robust against topic changes. We show the robustness of our features against topic changes by evaluating the features under increasing topic disjoint property of training and test documents. These experiments substantiate our claim that the features we propose capture the stylistic traits of authors that persist across multiple domains. The simplicity of our features also makes it scalable and hence can be applied to large scale data.

## References

- R. Arun, V. Suresh, and C.E.V. Madhavan. 2009. Stopword graphs and authorship attribution in text corpora. In *Semantic Computing, 2009. ICSC '09. IEEE International Conference on*, pages 192–196.
- John Blitzer, Ryan McDonald, and Fernando Pereira. 2006. Domain adaptation with structural correspondence learning. In *Conference on Empirical Methods in Natural Language Processing*, Sydney, Australia.
- John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jenn Wortman. 2008. Learning bounds for domain adaptation. In *Advances in Neural Information Processing Systems 21*, Cambridge, MA. MIT Press.
- Antonio Miranda Garca and Javier Calle Martn. 2007. Function words in authorship attribution studies. *Literary and Linguistic Computing*, 22(1):49–66.
- Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The weka data mining software: an update. *SIGKDD Explor. Newsl.*, 11:10–18, November.
- Dan Klein and Christopher D. Manning. 2003. A parsing: fast exact viterbi parse selection. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1, NAACL '03*, pages 40–47, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Kim Luyckx and Walter Daelemans. 2008. Authorship attribution and verification with many authors and limited data. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pages 513–520, Manchester, UK, August. Coling 2008 Organizing Committee.
- George Mikros and Eleni K. Argiri. 2007. Investigating topic influence in authorship attribution. In *PAN*.
- Frederick Mosteller and David L. Wallace. 1984. *Applied Bayesian and Classical Inference: The Case of the Federalist Papers*. Springer-Verlag.
- Fuchun Peng, Dale Schuurmans, Shaojun Wang, and Vlado Keselj. 2003. Language independent authorship attribution using character level language models. In *Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics - Volume 1, EACL '03*, pages 267–274, Stroudsburg, PA, USA.
- M. F. Porter, 1997. *An algorithm for suffix stripping*, pages 313–316. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- Sindhu Raghavan, Adriana Kovashka, and Raymond Mooney. 2010. Authorship attribution using probabilistic context-free grammars. In *Proceedings of the ACL 2010 Conference Short Papers, ACLShort '10*, pages 38–42, Stroudsburg, PA, USA.
- E. Stamatatos, N. Fakotakis, and G. Kokkinakis. 1999. Automatic authorship attribution. In *Proceedings of the ninth conference on European chapter of the Association for Computational Linguistics*, EACL '99, pages 158–164, Stroudsburg, PA, USA.
- Efstathios Stamatatos. 2009. A survey of modern authorship attribution methods. *J. Am. Soc. Inf. Sci. Technol.*, 60:538–556, March.
- Filiz Türkoğlu, Banu Diri, and M. Fatih Amasyali. 2007. Author attribution of turkish texts by feature mining. In *Proceedings of the intelligent computing 3rd international conference on Advanced intelligent computing theories and applications, ICIC'07*, pages 1086–1093, Berlin, Heidelberg. Springer-Verlag.