Report from the PAN'12 Evaluation Lab at CLEF'12

Benno Stein, Martin Potthast, Matthias Hagen, Tim Gollub, and Maik Anderka
Bauhaus-Universität Weimar
99432 Weimar, Germany
www.webis.de

Paolo Rosso, Alberto Barrón-Cedeño, and Parth Gupta
Universidad Politécnica de Valencia
Camino de Vera s/n. 46022 Valencia, Spain
http://users.dsic.upv.es/grupos/nle

Efstathios Stamatatos
University of the Aegean
83200 - Karlovassi, Samos, Greece
www.icsd.aegean.gr/lecturers/stamatatos

Moshe Koppel
Bar-Ilan University
52900 Ramat-Gan, Israel
http://u.cs.biu.ac.il/~koppel

Shlomo Argamon
Illinois Institute of Technology
10 W. 31st Street, Chicago, IL 60616, USA
http://lingcog.iit.edu/~argamon

Patrick Juola
Duquesne University
600 Forbes Avenue, Pittsburgh, PA 15282, USA
www.mathcs.duq.edu/~juola

Fabio Crestani and Giacomo Inches
University of Lugano
Via Giuseppe Buffi 13, CH-6904 Lugano, Switzerland
www.ir.inf.unisi.ch

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

Abstract

We report on the outcomes of the PAN evaluation lab on uncovering plagiarism, authorship, and social software misuse, which was held in conjunction with the CLEF'12 conference. A total of 72 people registered for the lab, 22 of whom presented their work on solving one or more of the lab's evaluation tasks. Two keynotes speakers were invited to talk about cross-language plagiarism detection, and tools that may be of help with this task. Besides the public sessions, nine of the 14 members of PAN's organizing committee attended the lab and discussed about the future of PAN in a steering committee meeting. The lab featured lively discussion and a number of interesting insights.

Summary

PAN is a networking initiative that centers around the topics of plagiarism, authorship, and social software misuse. In particular, we focus on methods and algorithms to assist text forensics and to extract an author's personal traits from her writing. Starting as a workshop in 2007, PAN is organized as an evaluation competition since 2009. As of 2010, PAN has been hosted at the CLEF conference, which provides an ideal platform to promote our mission, and to attract a growing number of participants for our evaluation tasks.

This year, PAN featured the four evaluation tasks Wikipedia quality flaw prediction, plagiarism detection, sexual predator identification in chat logs, and authorship attribution. Altogether, 38 teams submitted runs, ten of whom to more than one of the tasks. After the evaluation concluded, a total of 36 notebook papers were submitted, each describing a participant's approach to a given task, 25 of which have been presented at the workshop. The workshop was held in the course of 5 sessions totaling 8 hours at the conference. In addition to that, a 2.5 hour steering committee meeting was organized among the attending PAN organizers. Throughout the workshop, participants and attendees engaged in discussions about approaches and evaluation methodology. Finally, one of the PAN sessions was dedicated to two keynote speakers, Ralf Steinberger and Roberto Navigli, who described technologies applicable to cross-language plagiarism detection. Tables 1-3 overview the program. In what follows, the aforementioned activities and the obtained results are described in detail.

Task: Wikipedia Quality Flaw Prediction

This task has been run for the first time at PAN, and 21 teams registered to participated, whereas only 3 of them submitted runs to be evaluated, while 2 also submitted a notebook paper and attended the workshop to present their work. In addition, the task organizers wrote an overview paper which overviews the approaches and evaluation results in detail [1].

The online encyclopedia Wikipedia is one of the largest and most popular user-generated knowledge sources on the Web. Some facts: Wikipedia contains articles from more than 280 languages, the English Wikipedia version contains about 4 million articles, the Wikipedia community involves more than 35 million registered editors, and wikipedia.org ranks among the top ten most visited Web sites.[1] Probably the biggest challenge for Wikipedia pertains to the quality of its articles, since the community of Wikipedia authors is heterogeneous and since contributions to Wikipedia are not reviewed by experts before their publication. Both the size and the dynamic nature of Wikipedia render a comprehensive manual quality assurance infeasible. Hence the automated prediction of quality flaws in Wikipedia is a relevant problem, and the research on and the development of respective prediction approaches are the main goals of this task. The results of the 1st International Competition on Quality Flaw Prediction in Wikipedia can be summarized as follows: three quality flaw classifiers have been developed, which employ a total of 105 features to quantify the ten most important quality flaws in the English Wikipedia. Two classifiers achieve promising performance for particular flaws. An important "by-product" of the competition is the first corpus of flawed Wikipedia articles, the PAN Wikipedia quality flaw corpus 2012 (PAN-WQF-12).

At the conference, participants and interested attendees engaged in a discussion about the nature of this classification task, whether it is a one-class classification task, and how such tasks can be approached best. Moreover, the best performing participant received an award of appreciation from the German chapter of the Wikimedia Foundation. Finally, despite the comparably small number of participants, we still consider this task a success because of the

[1] Wikimedia, `http://meta.wikimedia.org/wiki/List_of_Wikipedias`.
Alexa Internet, Inc., `http://www.alexa.com/siteinfo/wikipedia.org`.

quality of the submitted approaches. Regarding the high number of registrants who did not submit anything, it may be the case that the task was considered being too difficult. Nevertheless, we have observed a large number of downloads of our publications about predicting Wikipedia quality flaws, including the task overview paper.

Task: Plagiarism Detection

This task has been run for the fourth time at PAN, and 39 teams registered to participated, and 17 of them submitted runs to be evaluated. Eight then submitted a notebook paper, seven of whom attended the workshop to present their work. In addition, the task organizers wrote an overview paper which overviews the approaches and evaluation results in detail [5].

Text plagiarism is perhaps one of the oldest forms of plagiarism, which, to this day, remains difficult to be identified in practice. Therefore, a lot of research has been conducted to detect plagiarism automatically. However, much of this research lacks proper evaluation, rendering it irreproducible at times and mostly incomparable across papers. In order to alleviate these issues, we have been organizing annual competitions on plagiarism detection since 2009, which included the development of the first standardized evaluation framework. In this year's fourth edition of the plagiarism detection competition, however, we again venture off the beaten track in order to further push the limits of evaluating plagiarism detectors. Our goal is to create a more realistic evaluation framework compared to our previous one in the course of the coming years. For this year's competition, we decided to construct a new plagiarism corpus comprising long, manually written documents. The corpus construction, for the first time, emulates the entire process of plagiarizing or reusing text, both at scale and in a controlled environment. The corpus comprises a number of features that set it apart from previous ones: (1) the topics of each plagiarized document in the corpus are derived from the topics of the TREC Web Track, and sources have been retrieved from the ClueWeb corpus. (2) The search for sources is logged, including click-through and browsing data. (3) A fine-grained edit history has been recorded for each plagiarized document. (4) A total of 300 plagiarized documents were produced, most of them 5000 words long, while ensuring diversity via crowdsourcing. Moreover, we asked participants to submit their detection software instead of just detection results on a given data set so as to analyze their approaches at our site. This allows for runtime comparisons, the use of confidential data sets for evaluation, and it enables us to compare performances using also our previous evaluation framework. The latter in particular solves the problem of incomparable evaluation results across different editions of PAN due ongoing improvements of the evaluation corpora. This year, most of the teams submitted their software, so that software submissions are apparently no big obstacle to participants and should be pursued further in the future.

At the conference, participants and interested attendees discussed about the how plagiarism detection performance should be measured and whether the existing measures are suitable and sufficient for the task. It was agreed they are, though there are still more performance aspects worthy measuring in order to better understand a plagiarism detector's performance. Moreover, the suitability of software submissions in general was discussed and whether the tools provided are sufficient for the task. Despite the fact there are still a number of improvements necessary for the tools to be of general use, we argued that an evaluation framework requires development time so as to study which of its parts can be improved, and how. This will be our focus of interest in the coming years.

Keynotes on Cross-language Plagiarism Detection

After the session on plagiarism detection, two keynotes and a panel discussion followed, focusing on the topic of cross-language plagiarism detection. We invited two well-known researchers

from the domain of cross-language information retrieval and natural language processing to give lectures on tools that may be of use in this regard, namely Ralf Steinberger from the European Commission's Joint Research Centre (JRC), Italy, and Roberto Navigli from the Univeristy La Sapienza, Italy.

A system that recognizes cross-lingual plagiarism needs to establish—among other things—whether two pieces of text written in different languages are equivalent to each other. Ralf Steinberger's keynote hence focused on this problem and the technology the Joint Research Centre (JRC) has been working on for many year: (a) cross-lingual document similarity calculation, (b) subject domain profiling of documents in many different languages according to the same multilingual subject domain categorization scheme, and (c) the recognition of name spelling variants for the same entity, both within the same language and across different languages and scripts). Steinberger's keynote covered further details on the JRC's motivation to work on multilingual and cross-lingual language technology applications, which is to offer readers and news analysts access to up to 150,000 online media articles per day in about fifty languages that are gathered and processed by the fully automatic Europe Media Monitor (EMM) family of applications. Moreover, he referred to freely available tools and data, such as the JRC EuroVoc Indexer, JEX, that can be used to represent documents in 22 different languages as lists of EuroVoc subject domain codes, and a number of further multilingual linguistic resources are available from `http://langtech.jrc.ec.europa.eu/JRC_Resources.html`.

Roberto Navigli's keynote was divided into two parts: in the first part of the talk, he presented BabelNet, a very large, wide-coverage multilingual semantic network. The resource is automatically constructed by means of a methodology that integrates lexicographic and encyclopedic knowledge from WordNet and Wikipedia. In addition Machine Translation is also applied to enrich the knowledge resource with lexical information for all languages. Experiments on new and existing gold-standard datasets have shown the high quality and coverage of the resource. Moreover, when provided with a vast amount of high-quality semantic relations, knowledge-rich word sense disambiguation algorithms compete with state-of-the-art supervised WSD systems in a coarse-grained all-words setting and outperform them on gold-standard domain-specific datasets. In the second part of the talk, Navigili analyzes cases in which BabelNet can be of help in cross-language plagiarism detection. He shows examples of how a large multilingual semantic network can provide hints for detecting plagiarized text.

Task: Sexual Predator Identification

This task has been run for the first time at PAN, and 16 teams submitted runs to be evaluated, twelve of whom also submitted a notebook paper and attended the workshop to present their work. In addition, the task organizers wrote an overview paper which overviews the approaches and evaluation results in detail [3].

"Chat messages" or "online conversations" are part of almost everybody's everyday life with services like Skype, Yahoo Messanger, MSN Messenger, ICQ but also IRC networks like Freenode or Quakenet. Although these services facilitate the establishment of new connections between persons or reinforce existing ones, they also allow for misbehaviors or cybercriminal acts. We hence organized the first international competition on Sexual Predator Identification and aimed at providing researches with a common framework to test methods for identifying such misbehaviors or cybercryminal activities. For simplicity, in the competition we only concentrate on the identification of "sexual predator" inside a chat, not dealing with other kind of misbehavior or media. A "sexual predator" is defined in the New Oxford American Dictionary as "a person or group that ruthlessly exploits others" while Wikipedia noticed how the definition "is used pejoratively to describe a person seen as obtaining or trying to obtain sexual contact with another person in a metaphorically "predatory manner". We refer to these interpretations of the term

"sexual predator" for the competition. In defining the tasks for the competition, we were also inspired by some previous works that addressed similar problem, even if none of them aimed at being an evaluation laboratory or containing a challenging collection to be used as a reference. In fact, we were the firsts to propose the following two kind of problems: given a collection containing chat logs involving two (or more) persons the participants had to (1) identify the predators among all users in the different conversations (problem 1) (2) identify the part (the lines) of the conversations which are the most distinctive of the predator behavior (problem 2). Given a realistic and challenging collection containing chat logs involving two (or more) persons, the 16 participants to the competition had to identify the predators among all the users in the different conversations and identify the part (the lines) of the predator conversations which were the most distinctive of the predator bad behavior. For the first problem we can conclude that lexical and behavioral features should be used when dealing with this kind of task. However, there is no unique method to identify predators but different approaches could be used, from SVM to Maximum-Entropy algorithm. Having a pre-filtering step to prune irrelevant conversations seems an important addition to the systems. For the second problem the most effective methods appeared to be those based on filtering on a dictionary or LM basis, partly due to the lack of ground truth for this specific problem (if we exclude the one based on 5-gram characters presence bit). The identification of common set of features and a group of effective strategies to identify predators is an achievement for this first part of the task. During the competition some issues were raised about the measurement of performances for the two problems, whether we should emphasize precision or recall and about the degree of subjectivity in the creation of the ground truth for problem 2. This is an achievement, too: with this competition we wanted to give researchers a unique place for comparing their methods but also for discussing and debating about future directions on this research area.

At the conference, discussions about this task centered around the definition of problem 2, and whether solving it is helpful to investigators. Moreover, a point of discussion was performance measurement and the weighting of precision and recall. A general consensus from the audience was that persecutors prefer recall over precision so as not to overlook offenders.

Task: Authorship Attribution

This task has been run for the second time at PAN, again attracting the large number of 12 teams submitted runs to be evaluated, all of whom also submitted a notebook paper and attended the workshop to present their work. In addition, the task organizers wrote an overview paper which overviews the approaches and evaluation results in detail [4].

Although, traditionally, authorship studies are done on the basis of close reading for stylistic detail, "nontraditional" or statistical authorship attribution has been around long enough to have developed into a traditional research problem of its own, especially in comparison to new tasks such as sexual predator identification. The task is well-understood (given a document, determine who wrote it) although amenable to many variations (given a document, determine a profile of the author; given a document pair, determine whether they were written by the same author; given a document, determine which parts of it were written my any specific person) and the motivation is clear. Applications for this technology include not only plagiarism detection but also historical inquiry, journalism, and legal dispute resolution (forensics). TREC-style competitive analyses of authorship methods using a standardized corpus have been around since at least 2004. This competition follows on the heels of a previous subtask at the PAN'11 competition, but differs from that competition in several ways: both the number and size of documents were decreased, the documents were taken from a different genre, the documents were no longer marked up extensively, a new sub-sub-task was added for authorship clustering. In future editions of this task, there are still plenty further issues to explore:

- Both this and the previous PAN competition have focused exclusively on English documents. Future competitions should include non-English languages.
- Similarly, different genre(s) should be represented, and the size of problems, including both number of authors and number and size of training/test documents should be varied.
- Some documents "written to order" should be included in order either to prevent cheating-by-Google and to ensure tighter control over the documents (especially for clustering/intrinsic plagiarism context).
- Professional forensic linguists should be invited to participate and compare their hand analyses with the results of the computer runs.
- Other types of (sub)tasks should be attached to the competitive study of authorship attribution, such as authorship profiling (e.g. was this document written in New York, London, or California? Was it written by a man or a woman?), document dating, and so forth.

These items also reflect most of the discussions about this task at the conference.

Steering Committee Meeting

The steering committee meeting at PAN was attended by nine of the 14 organizing committee members. For each of the aforementioned tasks at least one representative was able to make judgment calls and to discuss with the group the future of their respective tasks. Each task was discussed in turn, focusing on its current state, and what may still be achieved in the future. In the following, we give the main results of the discussion:

Wikipedia Quality Flaw Prediction. This task has attracted comparably little attention in terms of run submissions, though there was a high number of registrants. This may hint that the task is still too difficult in order to be accomplished during the competition's training and test phase. Given this fact, if the task is to be continued for another year, it should be done so in a more focused manner, so participants have less trouble to actually succeed at submitting the run. The high number of registrants as well as the large number of downloads of papers related to Wikipedia quality form our web pages suggest, that there is a big interest in this topic, which in turn would make a second edition of this task worthwhile. This, however, depends on whether the tasks main organizer, Maik Anderka, will be available for this in the next year. A possibility would to co-organize this task with one of the participants to lessen Anderka's workload.

Plagiarism Detection. This task has been revamped entirely this year, introducing a number of new tools and concepts, including a new corpus. The work that went into redesigning this task was devised from the start to last for several years, and in the coming years, the task will be reiterated using the new setup, while consolidating and improving it along the way. The feedback of the participants will be taken into account, and a number of new ideas for corpus construction have emerged which may be introduced in next years corpus release. Especially the introduction of software submissions instead of run submissions based on the TIRA experimentation platform allows for a number of interesting and insightful new analyses. However, TIRA will have to be developed further in order to streamline the evaluation process. The long-term goal for this is task is full automation.

Sexual Predator Identification. The introduction of the sexual predator identification task was a huge success this year. It was also covered on major news outlets such as Slashdot[2] which contributed a lot to its success in terms of participants. This interest, and the fact that defining

[2] http://science.slashdot.org/story/12/04/03/1734208

the task properly is still a work in progress (as per discussions at the workshop), suggest that the task be repeated at least once. Giacomo Inches, the main organizer of the task, however, mentions that repeating the task will only be possible if the workload is less than this year. This, in turn, raises the problem of obtaining a reasonable amount of training and test data for another edition, without relying on too big a manual effort. In this connection, it may be possible to generate appropriate test corpora automatically using tools such as ChatterBots or language models. Moreover, it may be possible to crowdsource the require data, or ask professional investigators whether they might share their data.

Authorship Attribution. The authorship attribution task continues to be a big interest, attracting many participants worldwide. The major problem of this task, too, is data acquisition, and in addition to that, the fact that the problem can be put in a large number of variants. Many participants complained about this fact at the workshop. Hence, it is suggested to focus the task further onto deciding whether two texts have the same author or no. For the problem of acquiring the appropriate data, there is still no appropriate solution, yet, limiting the task variants also reduces the work necessary to come up with a reasonable test corpus. New challenges for this task might include the introduction of texts in languages other than English, and the possibility to focus on the problem of author profiling (e.g., determining author demographics from their writing). Patrick Juola has agreed to again take charge of the next edition of this task.

PAN in general. For PAN in general, it was agreed that the number of tasks should not grow further but stay fixed at up to four. Rather, the quality of the tasks as well as the participants' contributions should be improved. Also, each task organizer should think about how to make contributions to the respective field of research by organizing their respective tasks. Moreover, given the success of software submissions, it was agreed to try and introduce them at all the tasks of PAN'13. This will help to further improve our tools already in place, and it will make evaluations reproducible in the future. With regard to organization, further advertisement of our tasks to new communities is adamant in order to broaden the community. Finally, more effort needs made to communicate with participants in a timely and unconfusing fashion.

Impact and Future of PAN

The impact of PAN can only be assessed in the long term, namely by checking whether the evaluation frameworks developed for PAN's tasks are picked up by the community in their daily research. Since this is only the fourth edition of PAN, this may not yet be possible, since most of its tasks have been introduced fairly recent. The task of plagiarism detection, however, has resulted in more than 165 references to the respective overview papers (according to Google Scholar). This shows that our efforts on organizing plagiarism detection evaluations has paid off. We expect similar results for the other tasks since they also are attractive to large communities of researchers, and since PAN is the only venue which focuses on their evaluation. This edition of PAN has had an reasonable impact, too, as could be observed at the conference. Despite PAN was run in parallel to to two other evaluation activities, the room was always filled with more than 30 people. The feedback received by the organizers with regard to PAN in general was very positive. In the future, we intend to make further contributions to the evaluation of each our tasks. This may also include the introduction of new, or the reformulation of long-standing problems in order to better pinpoint the challenging problems associated with them. Another important goal is sustainability: evaluations at PAN shall be developed to a point at which they can be run automatically. We positively wish to automate ourselves out of a job, thus freeing up time for new challenges.

Lab Program and Speakers

The following three tables overview the lab program of PAN, which is an excerpt of the CLEF conference program.

Table 1: PAN lab program on September 17.

Time	Session
14:00-16:00	**Lab overviews**
(15 min. talk)	PAN'12 - Uncovering Plagiarism, Authorship, and Social Software Misuse *Martin Potthast*
16:00-16:30	Coffee Break
16:30-17:00	**Poster Boaster Session**
17:00-18:30	**Poster Session**
	Encoplot - Tuned for High Recall (also proposing a new plagiarism detection score) - Notebook for PAN at CLEF 2012 *Cristian Grozea, Marius Popescu*
	A Set-Based Approach to Plagiarism Detection - Notebook for PAN at CLEF 2012 *Robin Küppers, Stefan Conrad*
	Applying Specific Clusterization and Fingerprint Density Distribution with Genetic Algorithm Overall Tuning in External Plagiarism Detection - Notebook for PAN at CLEF 2012 *Yurii Palkovskii, Alexei Belov*
	Detailed Comparison Module In CoReMo 1.9 Plagiarism Detector - Notebook for PAN at CLEF 2012 *Diego A. Rodríguez Torrejón, José Manuel Martín Ramos*
	Optimized Fuzzy Text Alignment for Plagiarism Detection - Notebook for PAN at CLEF 2012 *Fernando Sánchez-Vega, Manuel Montes-y-Gómez, Luis Villaseñor-Pineda*
	Bootstrapped Authorship Attribution in Compression Space - Notebook for PAN at CLEF 2012 *Ramon de Graaff, Cor J. Veenman*
	Paragraph Clustering for Intrinsic Plagiarism Detection using a Stylistic Vector-Space Model with Extrinsic Features - Notebook for PAN at CLEF 2012 *Julian Brooke, Graeme Hirst*
	Sub-Profiling by Linguistic Dimensions to Solve the Authorship Attribution Task - Notebook for PAN at CLEF 2012 *Upendra Sapkota, Thamar Solorio*
	Information Retrieval and Classification based Approaches for the Sexual Predator Identification - Notebook for PAN at CLEF 2012 *Darnes Vilariño, Esteban Castillo, David Pinto, Iván Olmos, Saul León*

Table 2: PAN lab program on September 19.

Time	Session
10:30-11:30	**Quality Flaw Prediction in Wikipedia**, Chair: Matthias Hagen
10:30-10:45	Overview of the 1st International Competition on Quality Flaw Prediction in Wikipedia *Maik Anderka, Benno Stein*
10:45-11:15	On the Use of PU Learning for Quality Flaw Prediction in Wikipedia - Notebook for PAN at CLEF 2012 *Edgardo Ferretti, Donato Hernández Fusilier, Rafael Guzmán Cabrera, Manuel Montes-y-Gómez, Marcelo Errecalde, Paolo Rosso*
11:15-11:30	FlawFinder: A Modular System for Predicting Quality Flaws in Wikipedia - Notebook for PAN at CLEF 2012 *Oliver Ferschke, Iryna Gurevych, Marc Rittberger*
11:30-12:30	**Plagiarism Detection**, Chair: Matthias Hagen
11:30-12:00	Overview of the 4th International Competition on Plagiarism Detection *Martin Potthast, Tim Gollub, Matthias Hagen, Jan Graßegger, Johannes Kiesel, Maximilian Michel, Arnd Oberländer, Martin Tippmann, Alberto Barrón-Cedeño, Parth Gupta, Paolo Rosso, Benno Stein*
12:00-12:15	Educated guesses and equality judgements: using search engines and pairwise match for external plagiarism detection - Notebook for PAN at CLEF 2012 *Lee Gillam, Neil Newbold, Neil Cooke*
12:15-12:30	Discussion
12:30-14:00	Lunch
14:00-14:15	**Plagiarism Detection**, Chair: Paolo Rosso Three way search engine queries with multi-feature document comparison for plagiarism detection - Notebook for PAN at CLEF 2012 *Simon Suchomel, Jan Kasprzak, and Michal Brandejs*
14:15-16:00	**Cross-Language Plagiarism Detection (Keynotes)**, Chair: Paolo Rosso
14:15-15:00	Cross-lingual Similarity Calculation for Plagiarism Detection and More - Tools and Resources *Ralf Steinberger*
15:00-15:45	Babelplagiarism: What can BabelNet do for Cross-language Plagiarism Detection? *Roberto Navigli*
15:45-16:00	Panel discussion
16:00-17:30	Break
17:30-20:00	**PAN Steering Committee Meeting**

Table 3: PAN lab program on September 20.

Time	Session
09:30-10:30	**Authorship Attribution**, Chair: Efstathios Stamatatos
09:30-10:00	An Overview of the Traditional Authorship Attribution Subtask + Mixture of Experts Authorship Attribution *Patrick Juola, Michael Ryan, and John Noecker Jr*
10:00-10:15	Authorship attribution: using rich linguistic features when training data is scarcen - Notebook for PAN at CLEF 2012 *Ludovic Tanguy, Franck Sajous, Basilio Calderone, Nabil Hathout*
10:15-10:30	Feature Bagging for Author Attribution - Notebook for PAN at CLEF 2012 *François-Marie Giraud, Thierry Artières*
10:30-11:00	Break
11:00-13:00	**Authorship Attribution & Sexual Predator Identification**, Chair: Patrick Juola
11:00-11:15	Graph-based and Lexical-Syntactic Approaches for the Authorship Attribution Task - Notebook for PAN at CLEF 2011 *Esteban Castillo, Darnes Vilariño, David Pinto, Iván Olmos, Jesús A. González, Maya Carrillos*
11:15-11:45	Overview of the International Sexual Predator Identification Competition at PAN-2012 *Giacomo Inches, Fabio Crestani*
11:45-12:00	Vote/Veto Classification, Ensemble Clustering and Sequence Classification for Author Identification - Notebook for PAN at CLEF 2012 *Roman Kern, Stefan Klampfl and Mario Zechner*
12:00-12:15	Quite Simple Approaches for Authorship Attribution, Intrinsic Plagiarism Detection and Sexual Predator Identification - Notebook for PAN at CLEF 2012 *Anna Vartapetiance, Lee Gillam*
12:15-12:30	Kernel Methods and String Kernels for Authorship Analysis - Notebook for PAN at CLEF 2012 *Marius Popescu, Cristian Grozea*
12:30-12:45	Conversation Level Constraints on Pedophile Detection in Chat Rooms - Notebook for PAN at CLEF 2012 *Claudia Peersman, Frederik Vaassen, Vincent Van Asch, Walter Daelemans*
12:45-13:00	Identifying Predators Using ChatCoder 2.0 - Notebook for PAN at CLEF 2012 *April Kontostathis, Will West, Andy Garron, Kelly Reynolds, Lynne Edwards*
13:00-14:00	Lunch
14:00-15:00	**Sexual Predator Identification**, Chair: Giacomo Inches
14:00-14:15	A Two-step Approach for Effective Detection of Misbehaving Users in Chats - Notebook for PAN at CLEF 2012 *Esaú Villatoro-Tello, Antonio Juárez-González, Hugo Jair Escalante, Manuel Montes-y-Gómez, and Luis Villaseñor-Pineda*
14:15-14:30	A Learning-Based Approach for the Identification of Sexual Predators in Chat Logs - Notebook for PAN at CLEF 2012 *Javier Parapar, David E. Losada, Alvaro Barreiro*
14:30-14:45	Features for modelling characteristics of conversations - Notebook for PAN at CLEF 2012 *Gunnar Eriksson, Jussi Karlgren*
14:45-15:00	Identifying Sexual Predators by SVM Classification with Lexical and Behavioral Features - Notebook for PAN at CLEF 2012 *Colin Morris, Graeme Hirst*

References

[1] Maik Anderka and Benno Stein. Overview of the 1st International Competition on Quality Flaw Prediction in Wikipedia. In Pamela Forner, Jussi Karlgren, and Christa Womser-Hacker, editors, *CLEF 2012 Evaluation Labs and Workshop – Working Notes Papers*, September 2012. ISBN 978-88-904810-3-1. URL `http://www.clef-initiative.eu/publication/working-notes`.

[2] Pamela Forner, Jussi Karlgren, and Christa Womser-Hacker, editors. *CLEF 2012 Evaluation Labs and Workshop – Working Notes Papers, 17-20 September, Rome, Italy*, 2012. ISBN 978-88-904810-3-1. URL `http://www.clef-initiative.eu/publication/working-notes`.

[3] Giacomo Inches and Fabio Crestani. Overview of the International Sexual Predator Identification Competition at PAN-2012. In Forner et al. [2]. ISBN 978-88-904810-3-1. URL `http://www.clef-initiative.eu/publication/working-notes`.

[4] Patrick Juola. An Overview of the Traditional Authorship Attribution Subtask. In Forner et al. [2]. ISBN 978-88-904810-3-1. URL `http://www.clef-initiative.eu/publication/working-notes`.

[5] Martin Potthast, Tim Gollub, Matthias Hagen, Jan Graßegger, Johannes Kiesel, Maximilian Michel, Arnd Oberländer, Martin Tippmann, Alberto Barrón-Cedeño, Parth Gupta, Paolo Rosso, and Benno Stein. Overview of the 4th International Competition on Plagiarism Detection. In Pamela Forner, Jussi Karlgren, and Christa Womser-Hacker, editors, *CLEF 2012 Evaluation Labs and Workshop – Working Notes Papers*, September 2012. ISBN 978-88-904810-3-1. URL `http://www.clef-initiative.eu/publication/working-notes`.