Shallow Text Analysis and Machine Learning for Authorship Attribution

Kim Luyckx and Walter Daelemans

CNTS-Language Technology Group, University of Antwerp, Belgium

Abstract

Current advances in shallow parsing and machine learning allow us to use results from these fields in a methodology for Authorship Attribution. We report on experiments with a corpus that consists of newspaper articles about national current affairs by different journalists from the Belgian newspaper *De Standaard*. Because the documents are in a similar genre, register, and range of topics, token-based (e.g., sentence length) and lexical features (e.g., vocabulary richness) can be kept roughly constant over the different authors. This allows us to focus on the use of syntax-based features as possible predictors for an author's style, as well as on those token-based features that are predictive to author style more than to topic or register. These style characteristics are not under the author's conscious control and therefore good clues for Authorship Attribution. Machine Learning methods (TiMBL and the WEKA software package) are used to select informative combinations of syntactic, token-based and lexical features and to predict authorship of unseen documents. The combination of these features can be considered an implicit profile that characterizes the style of an author.

1 Introduction

We define Authorship Attribution as the automatic identification of the author of a text on the basis of linguistic features of the text. Applications of Authorship Attribution range from resolving discussions about disputed authorship to forensic linguistics. In this paper, we interpret Authorship Attribution as a text categorization problem. The detection of age, region and gender of the author are other possible applications that could be handled this way, but will not be discussed here.

Automatic Text Categorization (Sebastiani 2002, 2) is a text mining application that labels documents according to a set of predefined content categories. Applications of Text Categorization are numerous. The most important ones are document indexing, document filtering or routing, and the hierarchical categorization of web pages and web search engines. Similar techniques are also being used at sentence level rather than document level for word sense disambiguation. Most Text Categorization systems use a two-stage approach in which (i) automatic feature selection is achieved of features (mostly terms, but also possibly n-grams of terms, NPs, ...) that have high predictive value for the categories to be learned, and (ii) a machine learning approach is used to learn to categorize new documents by using the features selected in the first stage. To allow the selection of linguistic features rather than (n-grams of) terms, robust and accurate text analysis tools such as lemmatizers, part of speech taggers, chunkers etc., are necessary.

An application of this methodology to Authorship Attribution starts from a set of training documents (documents of which the author is known), automatically

extracts features that are informative for the identity of the author and trains a machine learning method that optimally uses these features to do the author attribution for new, previously unseen, documents. Researchers assume that all authors have specific style characteristics that are outside their conscious control. On the basis of those linguistic patterns and markers, the author of a document can be identified (Diederich, Kindermann, Leopold and Paass 2000, 1–2). Rather than designing specific linguistic markers by introspection and testing them by hand, we will use automatic techniques to extract them from text and to test their usefulness in autorship attribution. We will use automatic text analysis tools (a lemmatizer, tagger, and other shallow parser modules) to allow the automatic extraction of potentially relevant linguistic features and patterns.

1.1 Features

We distinguish between four types of features that have traditionally been proposed as being able to differentiate between authors: token-level features (e.g., word length, syllables, *n*-grams), syntax-based features (e.g., part-of-speech tags, rewrite rules), features based on vocabulary richness (e.g., type-token ratio, hapax legomena) and common word frequencies (Stamatatos, Fakotakis and Kokkinakis 2001a). Most studies in the field are based on word forms and their frequencies of occurrence. Studies in the 1950's already were based on token-level features because no powerful computers and robust text analysis software were available (Holmes 1994). But today there are still researchers who use this type of features because it is simple and effective for Authorship Attribution. Stamatatos, Fakotakis and Kokkinakis (2001b) criticise token-level features, although some of their experiments are based on them:

> It is not possible for such measures to lead to reliable results. Therefore, they can only be used as complement to other, more complicated features (Stamatatos et al. 2001b, 195).

Features based on vocabulary richness are more complicated and relevant to an author's style, but have been criticised because they tend to be highly dependent on text length and unstable for texts shorter than 1,000 words (Stamatatos, Fakotakis and Kokkinakis 1999, 162). Common word frequencies can be calculated easily, but selecting the most appropriate words requires some effort.

The contrast between content and function words is basic in Authorship Attribution studies. Authors writing about the same topics tend to use a similar set of content words. Still, those authors have a conscious or unconscious preference for certain other content words. Function words do not seem at first sight to be reliable style markers, since they are very frequent and occur in every text. Nevertheless, the use and frequency of function words is characteristic for authors. An advantage of function words is that they are not under the author's conscious control (Holmes 1994, 90-91). Syntax-based features have been suggested as a different, new, path for capturing style. Though the results are promising (cf. Baayen, Van Halteren and Tweedie (1996), Diederich et al. (2000), Khmelev and

Tweedie (2001), Kukushkina, Polikarpov and Khmelev (2001) and Stamatatos et al. (1999)), many researchers try to avoid this type of features because they are hard to compute.

Thanks to improvements in shallow text analysis, we can currently extract reliable syntax-based features. In this paper, we compare token-level, lexical and syntax-based features on a corpus of newspaper articles written by three (groups of) authors. Syntax-based features are extracted by means of the Memory-Based Shallow Parser (MBSP) (Daelemans, Bucholz and Veenstra 1999), which gives an incomplete parse of the input text. MBSP does four types of analysis: it tokenizes the input, performs a Part-of-speech (POS) analysis, looks for noun phrase, verb phrase and prepositional phrase chunks and detects the subject and object of the sentence. The output (of the English MBSP trained on the Wall Street Journal corpus) looks like this:

> [NP1*Subject* POS//NNP tags/NNS NP1*Subject*] [VP1 can/MD be/VB subdivided//VBN VP1] PNP [P into/IN P] [NP open/JJ and/CC NP] PNP [VP2 closed/VBD VP2] [NP2*Object* class/NN words/NNS NP2*Object*] ./.

1.2 Learning Methods

The other focus in Authorship Attribution lies on the classification techniques to be applied. Although there are many techniques for Authorship Attribution, the majority of the studies applies statistical techniques because they are easy to compute and because they are believed to offer an objective method. We will show that the combination of shallow parsing for the automatic construction of predictive features with standard machine learning methods for feature selection and categorization provides an effective methodology for the development of Authorship Attribution systems.

For the Machine Learning experiments, we used a variety of algorithms available in the Waikato Environment for Knowledge Analysis (WEKA)[1] software package (Witten and Frank 1999). We will only report on results obtained with the neural network, traditionally a good approach for working with numeric data, and also one of the WEKA-provided algorithms with which the best results were obtained in exploratory experiments (Luyckx 2004). We also report on experiments using the Tilburg Memory-Based Learner (TiMBL) (Daelemans, Zavrel, van der Sloot and van den Bosch 2004), which is a more advanced $k - nn$-algorithm than the one provided in WEKA. Memory-Based Learning has been proposed as a learning method with the right kind of bias for learning language processing problems because of its ability to learn from untypical or low-frequency events in training data (Daelemans and van den Bosch 2005).

[1] WEKA, The University of Waikato: http://www.cs.waikato.ac.nz/~ml/index.html

2 Data and Features

The corpus used for training and testing consists of four hundred articles taken from the online archive of the Belgian daily newspaper *De Standaard*[2]. The goal of the experiments was to differentiate between two authors writing about national current affairs. In order to focus on the usefulness of syntactic and token-based features for author rather than topic or register detection, we chose documents within the same range of topics and in the same genre so that there is little difference in vocabulary and register. In order to test the system's robustness, there is a third class of 'other authors'. That way, the system will not only be able to identify the article as written by Anja Otte (class A) or Bart Brinckman (class B) but also as *not* being written by an author from a third class. This O-class consists of articles by ten other authors writing about national current affairs and some collaborative articles by Anja Otte and Bart Brinkman. These may be interesting for later research on the attribution of authorship to articles written by two authors. Table 1 gives an overview of the structure of the training and test corpus.

Author class	Training corpus		Test corpus	
	# articles	# words	# articles	# words
A (Anja Otte)	100 articles	57,682	34 articles	20,739
B (Bart Brinckman)	100 articles	54,479	34 articles	25,684
O (The Others)	100 articles	62,531	32 articles	21,871

Table 1: Training and test corpus

2.1 Features

All features used in this research were automatically extracted using output of the Memory-Based Shallow Parser and the Rainbow system for statistical text classification[3]. We selected nine feature sets of which five are syntax-based. The choice for those specific features is based among others on suggestions made by Glover and Hirst (1995, 4) concerning features based on tagged text. Another feature set (viz. *read*) is based on token information, and we also have two lexical feature sets. Combinations of all features and of all features except the lexical ones are also represented in two separate feature sets. Below is an overview of the feature sets involved in our research:

- *pos*: the frequency distribution of parts-of-speech (POS)
- *verb_B*: the frequency distribution of basic verb forms
- *verb*: the frequency distribution of verb forms

[2]De Standaard online: http://www.destandaard.be
[3]Rainbow: http://www-2.cs.cmu.edu/ mccallum/bow/rainbow

- *pat_num*: the frequency distribution of specific Noun Phrase patterns
- *function*: the frequency distribution of the fourty most frequent function words
- *lex*: the frequency distribution of the twenty most informative words according to the *Rainbow* program
- *read*: the readability score
- *all*: a combination of all features
- *syntax*: a combination of all syntax-based features and the token-level feature *read*

2.1.1 Parts-of-speech

Because most lexical features are highly author and language dependent, rules inferred by Machine Learning classifiers cannot be generalised to other authors or other languages (Stamatatos et al. 1999, 159). Syntax-based features like parts-of-speech do not have this problem because they are not under the conscious control of the author. According to Glover and Hirst (1995, 4), the distribution of parts-of-speech is a possible feature for Authorship Attribution. A list of the POS tags in the feature set and their mean frequency per text in the three author classes can be found below (cf. Table 2):

POS tag	Explanation	Frequency		
		A-class	B-class	O-class
ADJ	adjectives	35	39	41
BW	adverbs	35	30	34
LET	punctuation	79	64	73
LID	articles	59	63	66
N	nouns	121	118	137
SPEC	proper nouns	24	23	20
TSW	interjections	0.3	0.1	0.14
TW	numerals	8	7	14
VG	conjunctions	20	18	25
VNW	pronouns	50	38	48
VZ	prepositions	66	68	78
WW	verbs	81	76	89

Table 2: List of POS tags and their average frequency per text

2.1.2 Verb forms

According to Glover and Hirst (1995, 4), verb forms are also plausible syntax-based style markers. In order to be able to investigate how much grammatical information is needed, we decided to construct separate feature sets for basic and specific verb forms. Kukushkina et al. (2001, 181) found that using detailed information about grammatical classes was less effective than using generalized or 'incomplete' grammatical classes. The basic verb forms used by MBSP are based on the tagset of the Spoken Dutch Corpus (CGN), which distinguishes six verb forms: main verb singular, main verb plural, main verb ending in -t, infinitive, past participle and present participle (Hoekstra, Moortgat, Schuurman and van der Wouden 2001, 84-85). MBSP gives extra information about these verb forms, so that we end up with seventeen different verb forms.

2.1.3 Noun Phrase patterns

Word-class patterns are syntax-based features that also were proposed in (Glover and Hirst 1995, 4). A first step in investigating whether they are good predictors, is to indicate which specific Noun Phrase patterns occur in our corpus. After that, we construct a document feature vector for the distribution of those patterns. A complex noun phrase like *het sluitstuk van het cipiersakkoord van eind mei* is analysed by MBSP as LID N VZ LID N VZ N N. Most complex noun phrases consist of NP patterns combined by prepositions (VZ) or conjunctions (VG). Therefore, we distinguish twelve frequent NP patterns (cf. Table 3):

Pattern	Example
N, VNW or SPEC	mensen, hij, Albert
ADJ N	snel akkoord
LID N	de regering
N SPEC	voorzitter Verhofstadt
VNW N	zijn partij
TW N	zes maanden
LID ADJ N	de beste kandidaten
N ADJ N	eind vorige week
TW ADJ N	twee overwerkte politici
LID TW N	de vier zwaargewichten
N TW N	zondag 25 december
LID TW ADJ N	een derde nationale steekproef

Table 3: List of np patterns

2.1.4 Function words

The frequency distribution of the fourty most frequent function words in the corpus are represented in the *function* feature set. This allows us to test the relevance of

selecting function words as clues for Authorship Attribution.

2.1.5 Content words

This lexical feature set contains binary information about the 20 words with highest mutual information according to the *Rainbow* program for statistical text classification. Mutual Information (MI) is a feature selection method (Sebastiani 2002, 13). We use mutual information to determine which information is shared by the three author classes and which is able to distinguish between them. The 20 words with highest MI selected by Rainbow are *partij*, *SP.A*, *blijkt*, *zegt*, *wie*, *echter*, *altijd*, *aldus*, *evenwel*, *blok*, *VLD*, *beide*, *MR*, *gewest*, *tegelijk*, *steeds*, *erg*, *afgelopen*, *momenteel* en *wilde*.

2.1.6 Readability

The readability score is a statistical technique that computes readability based on the average number of syllables per word and the average number of words per sentence (i.e., the Flesch-Kincaid Readability Formula). We want to test whether authors writing for the same newspaper about similar topics have a similar readability[4].

2.1.7 Combination

We also combined the feature sets mentioned above in two separate feature sets: one containing all information and another one only containing token-level (viz. *read*) and syntax-based features. Stamatatos et al. (2001b, 195) state that token-level features alone cannot be useful for Authorship Attribution. They do believe that they can be reliable when used in combination with more complicated features. By combining feature sets, we can test this hypothesis.

3 Machine Learning Approach

Classification of a specific text according to a number of author categories is done by means of Machine Learning. Per document, a feature vector is constructed, containing comma-separated binary or numeric features on the basis of the information described in Section 2.1 and a class label (A, B or O). During training, the Machine Learning algorithms use the information from the training corpus to generate a model by means of which the unseen test instances can be classified. We use the neural network (backprop) implementation of the WEKA software package, and the memory-based classifier TiMBL. In the remainder of this section we briefly discuss and motivate the Machine Learning algorithms we will report the results of.

Artificial Neural Networks consist of a network of units. The input units which represent features are weighted by the strength of their associated connections, and

[4]Rudolf Flesch: http://www.mang.canterbury.ac.nz/courseinfo/AcademicWriting/Flesch.htm

their sum is calculated by a unit receiving input. If the sum is higher than a specified threshold, the output unit fires. The connection weights are computed using the Multi-Layer Perceptron learning rule. Since every author class has its own profile, other sets of features will appear to be meaningful for the A-class than for the B- and C-classes. In our set-up, the neural network has different nodes referring to the three author classes. Classification is performed by checking each test instance against these class thresholds. In our experiments, a backpropagation neural network is used. Training runs through five hundred epochs but is terminated when the error rate increases twenty times in a row. The momentum and the learning rate were fixed at 0.2 and 0.3, respectively.

TiMBL (Memory-based learning) is a supervised inductive algorithm for learning classification tasks based on the $k - nn$ algorithm with various extensions for dealing with nominal features and feature relevance weighting. Memory-based learning stores feature representations of training instances in memory without abstraction and classifies new (test) instances by matching their feature representation to all instances in memory, finding the most similar instances. From these "nearest neighbors", the class of the test item is extrapolated. See Daelemans et al. (2004) for a detailed description of the algorithms and metrics used in our experiments. All memory-based learning experiments were done with the TiMBL software package[5]. In order not to bias the comparison with neural networks (which were used "off the shelf"), we did no extensive model selection (optimization) of the parameters for TiMBL, but we selected the 10 nearest neighbours and added weights using the Information Gain metric.

4 Results

In this Section, we report results with the selected algorithms on the held-out test data. We report on experiments with three (A, B and O) author classes, and compare the use of TiMBL and neural networks for Authorship Attribution.

4.1 Neural Networks

Table 4 gives the results obtained with Neural Networks.

Pos is the best performing syntax-based feature set, with 50.6% F-score. A combination of all syntax-based features increases the F-score (viz., to 61.7%) and has least difficulties identifying the B-class. *Function* outperforms the syntax-based features with 2% in F-score. Combining all features allows the classifier to achieve an F-score of 71.3%, with a highest score on the B-class. For the three-author problem, we see that syntax-based features are able to compete with lexical features but that a combination of syntax-based and lexical features performs best.

[5]Available from http://ilk.uvt.nl

Data sets	Author classes			Average
	A-class	B-class	O-class	
pos	34.0%	56.8%	**61.0%**	**50.6%**
verb_B	**45.9%**	43.3%	45.5%	44.9%
verb	**59.3%**	49.1%	41.9%	50.1%
pat_num	48.6%	50.7%	**50.9%**	50.1%
function	**66.7%**	65.7%	59.4%	**63.9%**
lex	54.2%	**71.4%**	53.5%	59.7%
read	**61.1%**	57.5%	25.0%	47.9%
all	70.2%	**74.6%**	69.0%	**71.3%**
syntax	62.1%	**72.3%**	50.8%	**61.7%**

Table 4: Performance on three author classes by Neural Networks in WEKA

4.2 TiMBL

Table 5 represents results obtained with the memory-based learner TiMBL on three author classes. Considering the F-scores per author does not lead to coherent conclusions. The best syntax-based feature set is *pos*, with 47.7% F-score, while the lexical feature set *function* achieves 54.8%, outperforming the *lex* feature set consisting of content words. Combining all syntax-based features leads to a similar performance (57.3%), while a combination of all features achieves a mean F-score of 72.6%. We see that our syntax-based features achieve better than the *function* lexical feature set. TiMBL performs slightly better on a combination of all features than Neural Networks, but worse on the combination of syntax-based and token-level features.

Data sets	Author classes			Average
	A-class	B-class	O-class	
pos	43.3%	**54.9%**	44.9%	**47.7%**
verb_B	**53.8%**	43.8%	27.6%	41.7%
verb	43.6%	**46.9%**	34.5%	41.7%
pat_num	**53.2%**	50.0%	35.6%	46.3%
function	**65.7%**	55.7%	43.1%	**54.8%**
lex	44.4%	**59.4%**	51.2%	51.7%
read	**62.9%**	53.3%	36.4%	50.9%
all	**77.6%**	74.7%	65.5%	**72.6%**
syntax	59.4%	**61.7%**	50.9 %	**57.3%**

Table 5: Performance on three author classes by Timbl

5 Conclusions

In this paper we proposed a methodology for Authorship Attribution based on the combination of shallow parsing techniques for the extraction of linguistic features with machine learning techniques for feature weighting and author prediction. We illustrated the feasibility of the approach on a corpus consisting of newspaper articles about national current affairs written by three author groups. The linguistic features were computed using the Memory-Based Shallow Parser and Rainbow software packages. We experimented with a multi-class set-up in which the two target authors (categories A and B) had to be identified in a collection of documents in which some documents written by others were present as well (category O).

We compared the performance of a Neural Network (part of the WEKA machine learning software package) and a memory-based learner (TiMBL) for the problem. We found that the three classes can be identified by the Neural Network with an F-score of 71.3% by a combination of token-level, syntax-based and lexical features. Combining syntax-based features leads to an F-score comparable with that of a feature set consisting of the frequency distribution of function words. With a 72.6% F-score, TiMBL does slightly better. Syntax-based features even outperform lexical ones with TiMBL. Combining syntax-based and token-level features performs almost equally well as or even better than using a lexical feature set. The best syntax-based feature sets are based on the distribution of parts-of-speech. In most cases, the lexical feature consisting of function words works best for the newspaper articles in our corpus. Combining all syntax-based features increases the F-score considerably.

Direct comparison with previous other approaches is impossible, so the following overview of results in related research is of course only indicative. Frequencies of rewrite rules have been shown to be able to distinguish between authors, register and text type in 95% of the documents. Nevertheless, Baayen et al. (1996) point out that their method is too extensive to be used in actual Authorship Attribution practice:

> With the general lack of syntactically annotated text material, it is unlikely that the works in question are available in such an annotated form. (Baayen et al. 1996, 129)

Experiments using frequencies of word forms, word lengths, tagwords and bigrams of tagwords reported on by Diederich et al. (2000) obtained results between 60 and 80 percent. Recall values ranged between 55 and 100 percent for lexical features and between 15 and 40 percent for a combination of token-level and syntax-based features. This shows that our own test results are within line and even considerably better as far as syntax-based features are concerned. Cluster analysis, a statistical technique, can also be applied in Authorship Attribution. On third-person narratives only, frequencies of high-frequency words reach a 87.5% accuracy in work by Hoover (2001, 428). The success rate of Markov chains reaches 83.7% (Khmelev and Tweedie 2001, 306).

Stamatatos et al. (1999) extracted token-level, phrase-level and analysis-level features by means of the Sentence and Chunk Boundaries Detector (SCBD) and found an average error rate of 31% over all authors - and thus a success rate of 69% (Stamatatos et al. 1999, 162). Our combination of lexical, token-level and syntax-based features achieves 72.6% accuracy on the task, which is close to results of the above mentioned studies.

We conclude from our results that the syntax-based, lexical and token-level features we extracted are able to successfully tackle Authorship Attribution problems. As a matter of fact, a combination of our syntax-based features performs almost equally well as and in some cases even better than lexical and token-level features, which were believed to be the most reliable discriminators for authors.

6 Further research

We consider the experiments described here as an explorative study. The results obtained will be compared with methods more common in stylometrics, and the method has to be tested on several other types of text. There is a general worry with newspapers that the texts of the authors are often changed by editor(s).[6]

However, we believe the results clearly open up new perspectives for further research on combining automatically extracted syntax-based features and Machine Learning techniques for Authorship Attribution. More research will be done on syntax-based features based on parsed text, e.g. the frequency of clause types, syntactic parallelism and the ratio of main to subordinate clauses (Glover and Hirst 1995, 4). We will also explore the different applications of Authorship Attribution, like plagiarism detection and the detection of gender, region, and other properties of the author. Finally, we are currently also using syntax-based features and Machine Learning in a study on Middle-Dutch sermons in order to extract stylistic characteristics.

References

Baayen, H., Van Halteren, H. and Tweedie, F.(1996), Outside the cave of shadows: Using syntactic annotation to enhance authorship attribution, *Literary and Linguistic Computing* **11**(3), 121–131.

Daelemans, W. and van den Bosch, A.(2005), *Memory-Based Language Processing*, Cambridge, UK: Cambridge University Press.

Daelemans, W., Bucholz, S. and Veenstra, J.(1999), Memory-Based Shallow Parsing, *Proceedings of CoNLL-99*, pp. 53–60.

Daelemans, W., Zavrel, J., van der Sloot, K. and van den Bosch, A.(2004), TiMBL: Tilburg Memory Based Learner, version 5.1, reference guide, *Technical Report ILK Research Group Technical Report Series no. 04-02, 2004*, ILK Research Group, University of Tilburg.

Diederich, J., Kindermann, J., Leopold, E. and Paass, G.(2000), Authorship Attri-

[6] Many thanks to the anonymous CLIN reviewer for these and other usful remarks.

bution with Support Vector Machines, *Applied Intelligence* **19**(1-2), 109–123.

Glover, A. and Hirst, G.(1995), Detecting stylistic inconsistencies in collaborative writing, *Writers at work: Professional writing in the computerized environment*.

Hoekstra, H., Moortgat, M., Schuurman, I. and van der Wouden, T.(2001), Syntactic annotation for the Spoken Dutch Corpus project, *Proceedings of the Twelfth Meeting of Computational Linguistics in the Netherlands*, pp. 73–87.

Holmes, D.(1994), Authorship Attribution, *Computers and the Humanities* **28**(2), 87–106.

Hoover, D.(2001), Statistical stylistics and Authorship Attribution: An empirical investigation, *Literary and Linguistic Computing* **6**(4), 421–444.

Khmelev, D. and Tweedie, F.(2001), Using Markov chains for identification of writers, *Literary and Linguistic Computing* **16**(4), 299–307.

Kukushkina, O., Polikarpov, A. and Khmelev, D.(2001), Using literal and grammatical statistics for Authorship Attribution, *Problemy Peredachi Informatsii* **37**(2), 96–108. Translated as Problems of Information Transmission.

Luyckx, K.(2004), *Syntax-based features and Machine Learning techniques for Authorship Attribution*, Master's thesis, University of Antwerp.

Sebastiani, F.(2002), Machine Learning in automated text categorization, *ACM Computing Surveys* **34**(1), 1–47.

Stamatatos, E., Fakotakis, N. and Kokkinakis, G.(1999), Automatic Authorship Attribution, *Proceedings of EACL 99*.

Stamatatos, E., Fakotakis, N. and Kokkinakis, G.(2001a), Automatic text categorization in terms of genre and author, *Computational Linguistics* **26**(4), 471–495.

Stamatatos, E., Fakotakis, N. and Kokkinakis, G.(2001b), Computer-based Authorship Attribution without lexical measures, *Computers and the Humanities* **35**(2), 193–214.

Witten, I. and Frank, E.(1999), *Data Mining: Practical Machine Learning tools with Java implementations*, San Fransisco: Morgan Kaufmann.