Biomedical Text Mining: State-of-the-Art, Open Problems and Future Challenges

Andreas Holzinger[1], Johannes Schantl[1], Miriam Schroettner[1],
Christin Seifert[2], and Karin Verspoor[3,4]

[1] Research Unit Human-Computer Interaction, Institute for Medical Informatics, Statistics and Documentation Medical University Graz, Austria
{a.holzinger,j.schantl,m.schroettner}@hci4all.at
[2] Chair of Media Informatics, University of Passau, Germany
christin.seifert@uni-passau.de
[3] Department of Computing & Information Systems, University of Melbourne, Australia
[4] Health and Biomedical Informatics Centre, University of Melbourne, Australia
karin.verspoor@unimelb.edu.au

Abstract. Text is a very important type of data within the biomedical domain. For example, patient records contain large amounts of text which has been entered in a non-standardized format, consequently posing a lot of challenges to processing of such data. For the clinical doctor the written text in the medical findings is still the basis for decision making – neither images nor multimedia data. However, the steadily increasing volumes of unstructured information need machine learning approaches for data mining, i.e. text mining. This paper provides a short, concise overview of some selected text mining methods, focusing on statistical methods, i.e. Latent Semantic Analysis, Probabilistic Latent Semantic Analysis, Latent Dirichlet Allocation, Hierarchical Latent Dirichlet Allocation, Principal Component Analysis, and Support Vector Machines, along with some examples from the biomedical domain. Finally, we provide some open problems and future challenges, particularly from the clinical domain, that we expect to stimulate future research.

Keywords: Text Mining, Natural Language Processing, Unstructured Information, Big Data, Knowledge Discovery, Statistical Models, Text Classification, LSA, PLSA, LDA, hLDA, PCA, SVM.

1 Introduction and Motivation

Medical doctors and biomedical researchers of today are confronted with increasingly large volumes of high-dimensional, heterogeneous and complex data from various sources, which pose substantial challenges to the computational sciences [1], [2]. The majority of this data, particularly in classical business enterprise hospital information systems, is unstructured information. It is often imprecisely called unstructured data, which is used in industry as a similar buzz word as "big data". In the clinical domain it is colloquially called *"free-text"*,

A. Holzinger, I. Jurisica (Eds.): Knowledge Discovery and Data Mining, LNCS 8401, pp. 271–300, 2014.

which should be more correctly defined as *non-standardized data* [3], [?]. However, this unstructured information is particularly important for decision making in clinical medicine, as it captures details, elaborations, and nuances that cannot be captured in discrete fields and predefined nomenclature [5]. All essential documents of the patient records contain large portions of complex text, which makes manual analysis very time-consuming and frequently practically impossible, hence computational approaches are indispensable [6]. Here it is essential to emphasize that for the clinical doctor the written text in the medical findings is still the basis for decision making – neither images nor multimedia data [7].

2 Glossary and Key Terms

Bag-of-Words: A representation of the content of a text in which individual words (or terms) and word (or term) counts are captured, but the linear (sequential) structure of the text is not maintained. Processing of linguistic structure is not possible in this representation.

Classification: Identification to which set of categories (sub-populations) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known. Supervised machine learning technique.

Clustering: Grouping a set of objects in such a way that objects in the same group (cluster) are more similar to each other than to those in other groups (clusters). Unsupervised machine learning technique.

Corpus: A collection of (text) documents. In a vector space model, each document can be mapped into a point (vector) in $\mathbb{R}^n$. For specific applications, documents or portions of documents may be associated with labels that can be used to train a model via machine learning.

Knowledge Discovery: Exploratory analysis and modelling of data and the organized process of identifying valid, novel, useful and understandable patterns from these data sets.

Machine Learning: Study of systems that can learn from data. A sub-field of computational learning theory, where agents learn when they change their behaviour in a way that makes them perform better in the future.

Metric space: A space is where a notion of distance between two elements (a metric) is defined. The distance function is required to satisfy several conditions, including positive definiteness and the triangle inequality.

Optimization: The selection of a best element (with regard to some criteria) from some set of available alternatives.

Text: Text is a general term for sequences of words. Text may be further structured into chapters, paragraphs and sentences as in books. Texts may contain some structure easily identifiable for humans, including linguistic structure. However, from the computer science point of view, text is unstructured information because the structure is not directly accessible for automatic processing.

Term: Units of text, often representing entities. Terms may be words, but also may be compound words or phrases in some application scenarios (e.g., "IG-9", "Rio de Janeiro", "Hodgkin's Lymphoma"). Relevant terms may be defined by a dictionary; the dictionary subsequently defines the dimension of the vector space for representation of documents.

Text Mining (TM): or Text Data Mining. The application of techniques from machine learning and computational statistics to find useful patterns in text data [8].

Vector Space Model (VSM): Representation of a set of documents D as vectors in a common vector space. Approach whose goal is to make objects comparable by establishing a similarity measure between pairs of documents by using vector algebra (Euclidean distance, cosine similarity etc.). Most common vector space is the n-dimensional vector space of real numbers $\mathbb{R}^n$.

2.1 Notation

Document set or corpus D, $d \in D$
Vocabulary/Dictionary: V vocabulary, $t \in V$ term, $|V|$ size of the vocabulary
$\mathbb{R}^n$ n-dimensional space of real numbers
N - number of documents
$\overrightarrow{d}$ - vector representation of document d

3 Computational Representation of Text

In this chapter, we will introduce a selection of text mining methods, with an emphasis on statistical methods that utilise a matrix-like representation of the input data. In the case of text this representation is called the Vector Space Model (VSM). The general processing steps from textual data to the vector-space representation is described in section 3, the VSM itself is introduced in section 3.2. Specific algorithms applied to this representation of text will be presented in section 4.

There are a large number of linguistic approaches to processing of biomedical text, known as *biomedical natural language processing* (BioNLP) methods. Such approaches make extensive use of linguistic information such as grammatical relations and word order, as well as semantic resources such as ontologies and controlled vocabularies. These methods have been demonstrated to be particularly effective for extraction of biomedical concepts, events and relations, where the linguistic structure of the text can be particularly important. There are a number of resources that explore BioNLP methods, including a short encyclopedia chapter [9] and two recently published books [10, 12]. We therefore point the reader to those resources for more detail on such methods.

3.1 The Text Processing Pipeline

Text processing in the biomedical domain applies the same pipeline as for general text processing, an overview of which can be found in [13]. A detailed analysis of the steps in the context of information retrieval is available in [14]. Figure 1 shows and overview of a typical text processing pipeline assumed for statistical modeling of text. In the following we present each step in detail focusing on the content of the documents (and not the meta data).

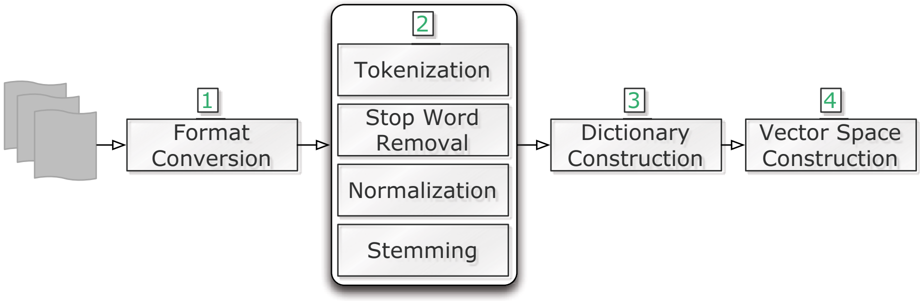

Fig. 1. Text processing pipeline. After converting documents of various source formats $\boxed{1}$, the terms for the dictionary are defined $\boxed{2}$, the dictionary is constructed $\boxed{3}$, which leads to the vector space representation for the documents $\boxed{4}$.

Format Conversion: Text comes in various formats, some of which can be accessed only by humans, such as paper books or lecture notes on a notepad. But even where text is digitally available, the formats vary and need to be normalized to apply text analysis methods. For instance, web pages contain not only the actual text content but also layout information (e.g., HTML tags, JavaScript code) and in HTML 5 the format has been defined with focus on layout and semantic structure [15]. On the other hand, Adobe's PDF format is optimized for layout and optimized printing [16]. In both HTML and PDF files, additional content, such as images or tables may be embedded. Source documents also come with various character encodings, like plain-text in Latin-1 encoding.

All such variants of texts have to be unified to make them accessible and the content available to text analysis methods. Plain text can be extracted from PDFs or Word documents, layout information is typically removed or ignored, and character encodings must be carefully handled, in particular to ensure correct treatment of special characters such as Greek letters or mathematical symbols. Depending on the source format this step may take considerable effort when the documents are in non-standard format. The output of the format conversion step are documents in a plain-text format with a standardized encoding suitable for further processing.

Tokenization: Vector space models typically use words as their basic representational element. Words must be identified from a text document, or a sequence of characters. This requires splitting those sequences into word-like pieces, a process called tokenization. In the simplest case tokenization involves splitting on non-alphanumeric characters (e.g., white spaces, punctuation marks, quotation marks). However, the process may be difficult in detail, e.g., using such simple splitting rules would split the words `O'Connor` and `isn't` in the same way with very different meaning. Further, specific tokenization rules may be necessary for some domains, e.g. by not splitting email addresses. The outcome of this step is a sequence of tokens for each document.

Stop Word Removal: Words that are thought not to carry any meaning for the purpose of text analysis are called "stop words". They are considered to be unnecessary and are typically removed before constructing a dictionary of relevant tokens. Removing those words serves two purposes: first, the reduced number of terms decreases the size of the vector space (reduced storage space) and second, the subsequent processes operating on the smaller space are more efficient. Stop words can be either removed by using predefined lists or applying a threshold to the frequency of words in the corpus and removing high-frequent words. Stop word removal is an optional step in text preprocessing. The outcome of this step is a sequence of tokens for each document (which may be smaller than the sequence obtained after tokenization).

Normalization: The process of normalization aims at finding a canonical form for words with the same semantic meaning but different character sequences. For instance, the character sequences `Visualization` and `visualisation`, and `IBM` and `I.B.M.` contain different characters but carry the same semantic meaning. Normalization methods include case-folding (convert all letters to lower case letters), and removing accents and diacritics, but their applicability depends of the language of the texts.

Stemming and Lemmatization: The underlying assumption for stemming and lemmatization is that for the purpose of machine text analysis the meaning of the words is not influenced by their grammatical form in the text. Stemming is the process of heuristically removing suffices from words to reduce the word to a common form (the word stem), while lemmatization refers to more sophisticated methods using vocabularies and morphological analysis. The BioLemmatizer [17] is a tool for lemmatization that is tailored to biomedical texts. The most common used stemming algorithm for the English language is the Porter stemmer [18]. The outcome of this step is the representation of documents as a sequence of (normalized, base) terms.

Building the Dictionary: Having obtained a sequence of terms for each document, a dictionary can be constructed from the document texts themselves. It is simply the set of all terms in all documents. Generally in machine learning one would refer to the dictionary terms as features. Details on how to obtain the vector space model of the document corpus given a dictionary and the documents are presented in section 3.2.

Note that in more linguistic approaches to text analysis, the dictionary may be an externally-specified collection of terms, potentially including multi-word terms, that reflect standard vocabulary terms for a given domain. For biomedicine, dictionaries of terms may include disease or drug names, or lists of known proteins. This dictionary may be used as a basis for identifying meaningful terms in the texts, rather than the more data-driven methodology we have introduced here. We will not consider this complementary approach further but again refer the reader to other resources [10, 12].

3.2 The Vector Space Model

The Vector space model aka term vector model is an algebraic model for representing any data objects in general, and text documents specifically, as vectors of so-called identifiers, e.g. index terms. The VSM is state-of-the-art in information retrieval, text classification and clustering for a long time [14, 19, 20]. In practice VSM's usually have the following properties, (1) a high dimensional features space, (2) few irrelevant features and (3) sparse instance vectors [21]. The VSM has been often combined with theoretical, structural and computational properties of connectionist networks in order to provide a natural environment for representing and retrieving information [22].

Functionality: Documents d in a corpus D are represented as vectors in a vector-space. The dimensionality of the vector space equals the number of terms in the dictionary V. The two most common vector space models used are the Boolean model and the real-valued vector-space. Figure 2 shows two example documents d_1 and d_2 in a three-dimensional vector space spanned by terms t_1, t_2 and t_3. In a metric vector space, similarities or distances between documents can be calculated, a necessary ingredient for most algorithms in information retrieval, text clustering and classification.

In a Boolean vector space, documents are representing as binary vectors, i.e. $d \in \{0, 1\}^{|V|}$. This model can only capture whether a certain term occurs in a document or not. On the contrary for documents in a real-valued vector space $d \in \mathbb{R}^{|V|}$ more information about the term-document relation can be captured by weighting the term occurrences. Several different ways of computing the term weights, have been developed. The most prominent weighting schemes are tf-idf weighting [19] and the BM-25 family of weighting schemes [23].

The intuition behind *tf-idf weighting* is that (i) terms that occur multiple times in a document should have a higher influence on the document (tf) and (i) terms that occur in many documents should have lower influence because they

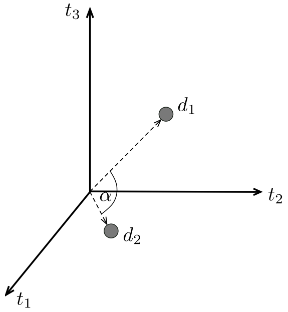

Fig. 2. A simple three-dimensional vector space with two documents. The angle between two documents can be used a similarity measure (cf. cosine similarity).

are less discriminating between documents (idf). More precisely, the $tf - idf$ weight of term t for document d is determined by

$$\text{tf-idf}_{t,d} = \text{tf}_{t,d} \cdot \text{idf}_t \quad (1)$$

$$\text{idf}_t = \log \frac{N}{\text{df}_t} \quad (2)$$

where $\text{tf-idf}_{t,d}$ is the number of occurrences of term t in document d, N is the number of documents and df_t the number of documents that contain term t.

In *BM-25 weighting* additional statistics of the text corpus, i.e. a document length-normalization, is incorporated in the weighting formula. For more details about BM-25 weighting variants see [23].

Cosine similarity The cosine similarity between document d_i and d_j corresponds to the angle between the documents (see figure 2 and is defined as

$$sim(d_i, d_j) = \frac{\overrightarrow{d_i} \cdot \overrightarrow{d_j}}{|\overrightarrow{d_i}||\overrightarrow{d_j}|} \quad (3)$$

where $\overrightarrow{d_i}, \overrightarrow{d_j}$ is the vector space representation and $|\overrightarrow{d_i}|, |\overrightarrow{d_j}|$ the Euclidean length of document d_i and d_j respectively. For normalized vectors the cosine similarity is equal to the dot product of those vectors, because the denominator in formula 3 becomes 1. Other similarity measure include Euclidean distance, which works similar well to cosine similarity for normalized document vectors and the dot product [14].

Examples in the Biomedical Domain: In Hlioautakis et al. (2006) [24] a good application example of the vector space model is explained. A medical knowledge finder which is based on the vector space model was evaluated in a clinical setting and compared with a gold standard by Hersh et al. [25]. In Müller et al. [26] content based image retrieval systems in the medical domain are compared with each other where some are based on vector space models.

According to Cunha et al. [27] a combination of vector space model, statistical physics and linguistics lead to a good hybrid approach for text summarization of medical articles.

Discussion: The VSM is the state-of-the art representation for text in information retrieval and text mining. *Advantages* of the model include:

- The model is simple and clear.
- The representation is a matrix that can be easily used by many machine learning algorithms.
- A continuous ranking of documents can be achieved according to their similarity to a query vector.
- A general representation of a document as a real-valued vector allows for different weighting schemes, which are used for instance to incorporate users' relevance feedback.

The *limitations* of the VSM are the following:

- The method is calculation intensive and needs a lot of processing time compared to a binary text representation. E.g., two passes over the document-term matrix are necessary for tf-idf weighting.
- The vector space for text has a high dimensionality (for natural language texts in the order of 10^4), and the matrix is a sparse matrix, since only a very small number of terms occur in each document. Thus, a sparse representation of the matrix is usually required to keep the memory consumption low.
- Adding new documents to the search space, means to recalculate/re-dimension the global term document matrix.
- Each document is seen as a bag of words, words are considered to be statistically independent. The meaning of the word sequence is not reflected in the model.
- Assumption a single term represents exactly one word sense, which is not true for natural language texts, which contain synonymous and polysemous words. Methods like word sense disambiguation have to applied in the preprocessing step.

Similar Spaces: The Semantic Vector Space Model (SVSM) is a text representation and searching technique based on the combination of Vector Space Model (VSM) with heuristic syntax parsing and distributed representation of semantic case structures. In this model, both documents and queries are represented as semantic matrices. A search mechanism is designed to compute the similarity between two semantic matrices to predict relevancy [28].

Latent semantic mapping (LSM) is a data-driven framework to model globally meaningful relationships implicit in large volumes of (often textual) data [29]. It is a generalization of latent semantic analysis.

4 Text Mining Methods

There are many different methods to deal with text, e.g. Latent Semantic Analysis (LSA), Probabilistic latent semantic analysis (PLSA), Latent Dirichlet allocation (LDA), Hierarchical Latent Dirichlet Allocation (hLDA), Semantic Vector Space Model (SVSM), Latent semantic mapping (LSM) and Principal component analysis (PCA) to name only a few.

4.1 Latent Semantic Analysis (LSA)

Latent Semantic Analysis (LSA) is both: a theory and a method for both extracting and representing the meaning of words in their contextual environment by application of statistical analysis to a large amount of text LSA is basically a general theory of acquired similarities and knowledge representations, originally developed to explain learning of words and psycholinguistic problems [30,31]. The general idea was to induce global knowledge indirectly from local co-occurrences in the representative text. Originally, LSA was used for explanation of textual learning of the English language at a comparable rate amongst schoolchildren. The most interesting issue is that LSA does not use any prior linguistic or perceptual similarity knowledge; i.e., it is based exclusively on a general mathematical learning method that achieves powerful inductive effects by extracting the right number of dimensions to represent both objects and contexts. The fundamental suggestion is that the aggregate of all words in contexts in which a given word does and does not appear provides a set of mutual constraints that largely determines the similarity of meaning of words and sets of words to each other. For the combination of Informatics and Psychology it is interesting to note that the adequacy of LSA's reflection of human knowledge has been established in a variety of ways [32]. For example, the scores overlap closely to those of humans on standard vocabulary and subject matter tests and interestingly it emulates human word sorting behaviour and category judgements [30]. Consequently, as a practical outcome, it can estimate passage coherence and the learnability of passages, and both the quality and quantity of knowledge contained in an textual passage (originally this were student essays).

Functionality: Latent Semantic Analysis (LSA) is primarily used as a technique for measuring the coherence of texts. By comparing the vectors for 2 adjoining segments of text in a high-dimensional semantic space, the method provides a characterization of the degree of semantic relatedness between the segments. LSA can be applied as an automated method that produces coherence predictions similar to propositional modelling, thus having potential as a psychological model of coherence effects in text comprehension [32].

Having t terms and d documents one can build a $t \times d$ matrix X. Often the terms within this matrix are weighted according to term frequency - inverse document frequency (fd-idf) [x]. The main method now is to apply the singular value decomposition (SVD) on X [y]. Therefore X can be disjointed into tree components $X = TSD^T$. T and D^T are orthonormal matrices with the eigenvectors of XX^T and X^TX respectively. S contains the roots of the eigenvalues of XX^T and X^TX.

Reducing the dimensionality can now be achieved by step-by-step eliminating the lowest eigenvalue with the corresponding eigenvectors to a certain value k. See relatedness to PCA (section 4.5).

A given Query q can now be projected into this space by applying the equation:

$$\frac{Q = q^T U_k}{diag(S_k)} \tag{4}$$

Having Q and the documents in the same semantic space a similarity measure can now be applied. Often used for example is the so called cosine similarity between a document in the semantic space and a query Q. Having two vectors A and B in the n dimensional space the cosine similarity is defined as:

$$cos(\phi) = \frac{A \cdot B}{\|A\| \, \|B\|} \tag{5}$$

Examples in the Biomedical Domain: Latent semantic analysis can be used to automatically grade clinical case summaries written by medical students and therefore proves to be very useful [33]. In [34] latent semantic analysis is used to extract clinical concepts from psychiatric narrative. According to [35] LSA was used to extract semantic word and semantic concepts for developing a ontology-based speech act identification in a bilingual dialogue system. Furthermore, latent semantic analysis combined with hidden Markov models lead to good results in topic segmentation and labelling in the clinical field [36]. In [37] the characteristics and usefulness of distributional semantics models (like LSA, PLSA, LDA) for clinical concept extraction are discussed and evaluated.

Discussion: The *advantages* of LSA are:

- LSA tends to solve the synonym problem [38].
- LSA reflects the semantic of the texts, so similar concepts are found.
- Finds latent classes.
- Reduction of dimensionality of Vector Space Model.

LSA has the following *limitations*:

- High mathematical complexity (SVD).
- Recalculation of the singular value decomposition when adding new documents or terms.
- Offers only a partial solution to the polsemy problem [38].
- The estimation of k, that means how many eigenvalues to keep to get good information retrieval results.
- It has a bad statistical foundation [39].

4.2 Probabilistic Latent Semantic Analysis (PLSA)

The probabilistic latent semantic analysis (PLSA) is a statistical method for factor analysis of binary and count data which is closely related to LSA, however, in contrast to LSA which stems from linear algebra and performs a Singular Value Decomposition of co-occurrence tables, the PLSA technique uses a generative latent class model to perform a probabilistic mixture decomposition [40,41]. This results in a more principled approach with a solid foundation in statistical inference. PLSA has many applications, most prominently in information retrieval, natural language processing, machine learning from text, see e.g, [42–45]).

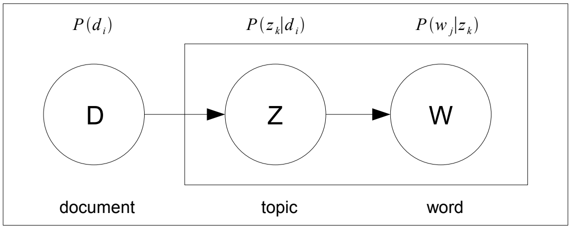

Fig. 3. Two level generative model

Functionality: The PLSA is a unsupervised technique that forms a two level generative model (see picture 3). In contrast to the LSA model, it builds up a clear probability model of the underlying documents, concepts (topics) and words and therefore is more easy to interpret. Abstractly spoken every document can talk about different concepts to a different extend as well as words form different topics. So having a document collection $D = \{d_1, d_2, d_i..., d_M\}$ we can define a co-occurrence table using a vocabulary $W = \{w_1, w_2, w_j..., w_N\}$. Furthermore each observation, where an observation is defined as occurrence of a word in a document is associated with an unobservable class variable $Z = \{z_1, z_2, z_k..., z_K\}$. We can now define a joint probability $P(d_i, z_k, w_j)$ as:

$$P(d_i, z_k, w_j) = P(w_j|z_k)P(z_k|d_i)P(d_i) \tag{6}$$

Getting rid of the unobservable variable z_k we can rewrite the equation as:

$$P(d_i, w_j) = \sum_{k=1}^{K} P(w_j|z_k)P(z_k|d_i)P(d_i) \tag{7}$$

The probability of the entire text corpus can now be rewritten as:

$$P(d, w) = \prod_{i=1}^{M} \prod_{j=1}^{N} P(d_i) \sum_{k=1}^{K} P(w_j|z_k)P(z_k|d_i) \tag{8}$$

Rewriting the probability of the text corpus in terms of a log likelihood yields:

$$L = log \prod_{i=1}^{M} \prod_{j=1}^{N} P(d_i, w_j)^{n(d_i, wj)} \tag{9}$$

Maximizing L is done by using the expectation maximization algorithm [46]. The algorithm basically consists of two steps:

Expectation-Step: Expectation for the latent variables are calculated given the observations by using the current estimates of the parameters.

Maximization-Step: Update the parameters such that L increases using the posterior probabilities in the E-Step.

The steps are repeated until the algorithm converges, resulting in the quantities $P(w_j|z_k)$ and $P(z_k|d_i)$.

Examples in the Biomedical Domain: According to [47] probabilistic latent semantic analysis is used to find and extract data about human genetic diseases and polymorphism. In [48] the usefulness of PLSA, LSA and other data mining method is discussed for applications in the medical field. Moreover PLSA counts to one of the well-known topic models that are used to explore health-related topics in online health communities [49]. According to Masseroli et al. [50] PLSA is also used for prediction of gene ontology annotations and provides good results.

Discussion: The *advantages* of PLSA include:

- Can deal with the synonym and polysemy problem [39].
- PLSA performs better than LSA [39].
- PLSA is based on a sound statistical foundation [39].
- Finds latent topics.
- Topics are easily interpretable.

PLSA has the following *limitations*:

- The underlying iterative algorithm of PlSA converges only logically [51].

4.3 Latent Dirichlet Allocation (LDA)

Latent Dirichlet allocation (LDA) is a generative probabilistic model for collections of discrete data such as text corpora, based on a three-level hierarchical Bayesian model, in which each item of a collection is modelled as a finite mixture over an underlying set of topics [52]. Each topic is, in turn, modelled as an infinite mixture over an underlying set of topic probabilities. In the context of text modelling, the topic probabilities provide an explicit representation of a document, consequently LDA can be seen as a 'bag-of-words' type of language modelling and dimension reduction method [53] (TODO: get pdf). One interesting application of LDA was fraud detection in the telecommunications industry in order to build user profile signatures and assumes that any significant unexplainable deviations from the normal activity of an individual end user is strongly correlated with fraudulent activity; thereby, the end user activity is represented as a probability distribution over call features which surmises the end user calling behaviour [54] (TODO: get pdf). LDA is often assumed to be better performing than e.g. LSA or PLSA [55].

Functionality: The basic idea behind LDA is that documents can be represented as a mixture of latent topics, which are represented by a distribution across words [52]. Similar to LSA and PLSA, the number of latent topics used in the model has to be fixed a-priori. But in contrast, for instance, to LSA which uses methods of linear algebra, LDA uses probabilistic methods for inferring the mixture of topics and words. According to Blei et al. [52], the assumed generative process for each word w in a document within a corpus D contains the following steps: (1) Choose N, the number of words for a document, estimated by a Poisson distribution. (2) Choose a topic mixture θ according to a Dirichlet distribution over α, a Dirichlet prior over all documents. (3) Each of the N words w_n are selected by first choosing a topic which is represented as multinomial random variable z, and second by choosing a word from $p(w_n|z_n, \beta)$, a multinomial probability conditioned on the topic z_n.

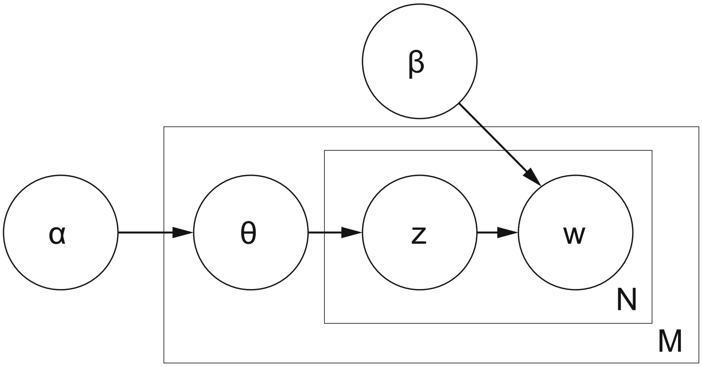

Fig. 4. LDA graphical model. The outer box represents documents, while the inner box represents the iterative choice of topics and words in a document. [52]

Figure 4, shows the probabilistic graphical model of LDA. The LDA representation contains three levels. The corpus level parameters α and β are sampled once within the process of generating a corpus. The document level variable θ is sampled once per document, and the word level parameters z_n and w_n are sampled once for each word within a document.

The goal of inference in the LDA model is to find the values of ϕ, the probability of word w occurring in topic z and θ, the distribution of topics over a document. There are several algorithms proposed for solving the inference problem for LDA, including a variational Expectation-Maximization algorithm [52], an Expectation-Propagation algorithm [56] or an collapsed Gibbs sampling algorithm of Griffiths and Steyvers [57].

Examples in the Biomedical Domain: In [58], Wu et al. apply LDA to predict protein-protein relationships from literature. One of their features to rank candidate gene-drug pairs was the topic distance derived from LDA. Case-based information retrieval from clinical documents due to LDA can be used to discover relevant clinical concepts and structuring in patient's health record [59]. In [60] [get pdf and check if LDA is used for topic model], Arnold and Speier present a topic model tailored to the clinical reporting environment which allows for individual patient timelines. In [61] unsupervised relation discovery with sense disambiguation is processed with the help of LDA which can be used in different areas of application. Moreover with LDA a genomic analysis of time-to-event outcomes can be processed [62]. According to [63] big data is available for research purposes due to data mining of electronic health records, but it has to pre-processed to be useful with the help of different mining techniques including LDA.

Discussion: LDA has the following *advantages*:

- Find latent topics.
- Topics are easy to interpret.
- Reduction of dimensionality of Vector Space Model.
- Better performance than e.g. LSA or PLSA [55].

The main *disadvantage* of LDA is the parameter estimation, e.g., the number of topics has to be known or needs to be estimated.

4.4 Hierarchical Latent Dirichlet Allocation (hLDA)

While topic models such as LDA treat topics as a flat set of probability distributions and can be used to recover a set of topics from a corpus, they can not give insight about the abstraction of a topic or how the topics are related. The hierarchical LDA (hLDA) model is a non-parametric generative probabilistic model and can be seen as an extension of LDA. One advantage of non-parametric models is, that these models do not assume a fixed set of parameters such as the number of topics, instead the number of parameters can grow as the corpus grows [64]. HLDA is build on the nested Chinese restaurant process (nCRP) to additionally organize topics as a hierarchy. HLDA arranges topics into a tree, in which more general topics should appear near the root and more specialized ones closer to the leaves. This model can be useful e.g. for text categorization, comparison, summarization, and language modelling for speech-recognition [64].

Functionality: HLDA uses the nCRP, a distribution on hierarchical partitions, as a non-parametric prior for the hierarchical extension to the LDA model [64]. A nested Chinese restaurant process can be illustrated by an infinite tree where each node denotes to a restaurant with an infinite number of tables. One restaurant is identified as the root restaurant, and on each of its tables lies a card with the name of another restaurant. Further, on each table of these restaurants are again cards to other restaurants and this structure repeats infinite times, with the restriction that each restaurant is referred only one time. In this way all restaurants are structured into an infinity deep and infinitely branched tree. Every tourist starts by selecting a table in the root restaurant according to the CRP distribution defined as:

$$\begin{aligned} p(occupied\ table\ i|previous\ customer) &= \frac{m_i}{\gamma + m - 1} \\ p(next\ occupied\ table|previous\ customer) &= \frac{\gamma}{\gamma + m - 1} \end{aligned} \tag{10}$$

where m_i denotes to the number of previous customers at table i, and γ is a parameter [64]. Next, the tourist chooses in the same way a table in the referred restaurant from the last restaurants table. This process repeats infinity many times. After M tourists go through this process the collection of paths describes a random sub-tree of the infinite tree [64].

According to [64], the nCRP is augmented in two ways to obtain a generative model for documents. (1) Each node in the tree is associated with a topic, which is a probability distribution across words. A path in the tree samples an infinite collection of topics. (2) The GEM distribution [65] is used to define a probability distribution on the topics along the path. A document is then generated by repeatedly sampling topics according to the probabilities defined by one draw of the GEM distribution, and further sampling each word from the probability distribution from the selected topic.

For detailed information about the inference in hLDA, it is referred to the original description of HLDA from Blei et.al [64].

Examples in the Biomedical Domain: Due to hLDA individual and population level traits are extracted from clinical temporal data and used to track physiological signals of premature infants and therefore gain clinical relevant insights [66]. According to [67] clinical document labelling and retail product categorization tasks on large-scale data can be performed by hLDA .

Discussion: Hierarchical LDA has the same advantages and limitations as LDA (see section 4.3), additionally *advantages* are:

- Organics topics in a hierarchy depending on their abstraction level.
- Number of topics does not need be fixed a-priori.

4.5 Principal Components Analysis

Principal component analysis (PCA) is a technique used to reduce multidimensional data sets to lower dimensions for analysis. Depending on the field of application, it is also named the discrete Karhunen-Loève transform, the Hotelling transform [68] or proper orthogonal decomposition (POD).

PCA was introduced in 1901 by Karl Pearson [69]. Now it is mostly used as a tool in exploratory data analysis and for making predictive models. PCA involves the calculation of the eigenvalue decomposition of a data covariance matrix or singular value decomposition of a data matrix, usually after mean centering the data for each attribute. The results of a PCA are usually discussed in terms of component scores and loadings.

Functionality: In this section we give a brief mathematical introduction into PCA, presuming mathematical knowledge about: standard deviation, covariance, eigenvectors and eigenvalues. Lets assume to have M observations of an object gathering at each observation N features. So at any observation point we can collect a N-dimensional feature vector Γ. All feature Vectors together form the $N \times M$ observation matrix $\Theta(\Gamma_1, \Gamma_2, ..., \Gamma_M)$. Furthermore the average feature vector is given by:

$$\Psi = \frac{1}{M} \sum_{n=1}^{M} \Gamma_n \tag{11}$$

So by mean-adjusting every feature vector Γ_i by $\Phi_i = \Gamma_i - \Psi$ one can form the covariance matrix of the mean adjusted data by:

$$C = \frac{1}{M} \sum_{n=1}^{M} \Phi_n \Phi_n^T \tag{12}$$

Basically the PCA is done by following steps:

- Calculate the eigenvalues and the corresponding eigenvectors of C, resulting in $\lambda_1, \lambda_2, ..., \lambda_M$ eigenvalues with the corresponding eigenvectors $u_1, u_2, ..., u_M$
- Keep the the highest eigenvalues $M' < M$ forming a matrix $P = [u_1, u_2, ..., u_{M'}]^T$ with the corresponding eigenvectors where $\lambda_1 > \lambda_2 > ...\lambda_{M'}$.
- Transform the data into the reduced space applying $Y = PC$

An important thing to mention is that $cov(Y) = \frac{1}{M'} YY^T$ is a Diagonalmatrix. That means we found a representation of the data with minimum redundancy and noise (off diagonal elements of the covariance matrix are zero). Remember that one diagonal element of the covariance matrix represent the variance of one typical feature measured. Another thing to mention for a better understanding is that the eigenvectors try to point toward the direction of greatest variance of the data and that all eigenvectors are orthonormal. See figure 5 for an illustrative example. PCA is closely related to SVD, so LSA is one of the typical examples in the field of information retrieval where PCA is used.

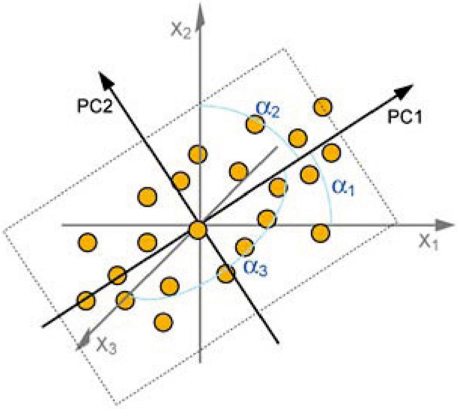

Fig. 5. Eigenvectors pointing to direction of maximum variance

Examples in the Biomedical Domain: Examples in medicine span a lot of different areas, due to the case that PCA is a base statistical mathematical tool. Especially used as space dimension reduction method, high dimensional parameter space can be reduced to a lower one. Nevertheless we want to give some recent research papers in various fields of medicine to show that even though the mathematical principles are know more than 100 years old [69], its appliance in modern medical informatics is still essential. According to [70] the principal components analysis is used to explicitly model ancestry differences between cases and controls to enable detection and correction of population stratification on a genome-wide scale. Moreover PCA is used in [71] to investigate a large cohort with the Tourette syndrome and evaluate the results. In [72] lay requests to gain medical information or advice from sick or healthy persons are automatically classified due to different text-mining methods including PCA. Furthermore, with the help of PCA and other text mining methods drug re-purposing can be identified trough analysing drugs, targets and clinical outcomes [73].

Discussion: PCA is widely used for dimensionality reduction and generally has the following *advantages*:

- Finds the mathematically optimal methods in the sense of minimizing the squared error.
- The measurement of the variance along each principle component provides a means for comparing the relative importance of each dimension.
- PCA is completely non-parametric.
- PCA provides the optimal reduced representation of the data.

However, the following *limitations* have to be considered when considering to apply PCA:

- Sensitive to outliers.
- Removing the outliers before applying the PCA can be a difficult task.
- Standard PCA can not capture higher order dependencies between the variables.

Similar Methods: Kernel PCA [74] is based on the kernel trick which comes from the field of Support Vector Machines but has successfully been applied to the PCA. Basically the kernel trick maps features that are not separable in the current space into a high dimensional where separation is possible. For this purpose one has basically just to know the Kernel function $K(x, y)$. A often used kernel function is the Gaussian kernel,

$$K(x, y) = e^{\frac{-||x-y||^2}{2\sigma^2}} \tag{13}$$

but other Kernels can also be used that at least that they have to fulfil the condition that the dot product between two components in feature space has the same results applying the kernel function between two components in the original space. The main advantage of kernel PCA is the fact that it is able to separate components, where normal PCA failures.

4.6 Classification of Text Using Support Vector Machines

Support Vector Machines are first introduced in the year 1995 by Vapnik and Cortes [75], and are one of the most commonly used classification methods. They are based on well founded computational learning theory and are analysed in research in very detail. SVMs can easily deal with sparse and high dimensional data sets, therefore, they fit very well for text mining applications [21].

Functionality: The task of a SVM is to find an hyperplane which separates the positive and negative instances (in case of a two class input data set) with the maximum margin. Therefore SVMs are also called maximum margin classifiers. The margin is defined as the distance between the hyperplane and the closest point to the hyperplane among both classes denoted by x_+ and x_-. Assuming that x_+ and x_- are equidistant to the hyperplane with a distance of 1, the margin is defined as:

$$m_D(f) = 1/2\hat{w}^T(x_+ - x_-) = \frac{1}{||w||} \tag{14}$$

where $\hat{w}$ is a unit vector in the direction of w which is known as the weight vector [76]. Maximizing the margin $\frac{1}{||w||}$, which is equivalent to minimizing $||w||^2$, will maximize the separability between the classes. In this context, one can distinguish between the hard margin SVM, which can be applied for linearly separably data sets, and the soft margin SVM which works also for not linearly separably data by allowing also misclassifications. The optimization task of a hard margin SVM is to minimize $1/2||w||^2$ subject to

$$y_i(w^T x_i + b) \geq 1, i = 1, ..., n \tag{15}$$

where y_i denotes the label of an example [76]. In case of the soft margin SVM the objective of minimizing $1/2||w||^2$. is augmented with the term $C\sum_{i=1}^{n}\xi_i$ to penalize misclassification and margin errors, where ξ_i are slack variables that allow an example to be in the margin $(0 \leq \xi_i \leq 1)$ or to be misclassified $(\xi_i > 1)$ and the parameter C sets the relative importance of minimizing the amount of slack and maximizing the margin [76]. The optimization task of a soft margin SVM is then to minimize $1/2||w||^2 + C\sum_{i=1}^{n}\xi_i$ subject to

$$y_i(w^T x_i + b) \geq 1 - \xi_i, x_i \geq 0 \tag{16}$$

The SVM algorithm as presented above is inherently a binary classifier. SVM has been successfully applied to multi-class classification problems by using the multiple one-vs-all or one-vs-one classifiers, a true multi-class SVM approach has been proposed by Crammer & Singer [77]. In many applications a non-linear classifier provides better accuracy than a linear one. SVMs use also the kernel trick, to transfer the input space into a higher dimensional space in which the data is linearly separable by a hyperplane. In this way SVMs fit also very well as a non-linear classifier.

Examples in the Biomedical Domain: According to Guyon et al. [78] a new method of gene selection based on support vector machines was proposed and evaluated. In a text classification task, Ghanem et al. [79], used SVMs in combination with regular expressions to evaluate whether papers from the Fly-Base data set should be curated based on the presence of evidence of Dosophila gene products. Support vector machine is also used for mining biomedical literature for protein-protein interactions Donaldson et al. [80]. To discover protein functional regions, Eskin and Agichtein [81] combined text and sequence analysis by using an SVM classifier. In Joshi et al. [82] support vector machines are explored for the use in the domain of medical text and compared with other machine learning algorithm. Further, SVMs are also used in the most effective approaches for the assertion checking and and relation extraction subtask in the i2b2 2011 challenge which are tasks of biomedical NLP [83]. Also in the i2b2 co-reference challenge 2012, SVM was used among the leading approaches [83].

Discussion: The SVM is considered state-of-the-art in text classification due to the following *advantages*:

- Can deal with many features (more than 10000).
- Can deal with sparse document vectors.
- Not restricted to linear models due to the kernel functions.

The only *disadvantage* of SVM classification is the interpretable of the classification model compared to easily interpretable models like Naive Bayes [84] or Decision trees [85].

5 Open Problems

5.1 Deployment and Evaluation of Text Mining in Clinical Context

Evaluation of text mining systems is substantially more mature in the context of analysis of the published biomedical literature than in the clinical context. A number of community evaluations, with shared tasks and shared data sets, have been performed to assess the performance of information extraction and information retrieval from the biomedical literature (e.g. the BioNLP shared tasks [86], BioCreative [87], and TREC Genomics [88]). Especially when dealing with supervised machine learning methods, one needs substantial quantities of labelled training data and again there are growing large-scale richly annotated training data resources (e.g. the GENIA [89] and CRAFT [90, 91] corpora, as well as a increasing number of more specific resources such as the SCAI chemical compounds corpus [92] and the Variome genetic variants corpus [93]).

Such tasks and resources have been far more limited on the clinical side. The 2012 TREC Medical Records track [94], i2b2 natural language processing shared tasks addressing clinical discharge summaries [95, 96], and the ShARE/CLEF eHealth evaluation lab 2013 [97] provide examples of the importance of such evaluations in spurring research into new methods development on real-world problems. The availability of de-identified clinical records that can be shared publicly is critical to this work, but has posed a stumbling block to open assessment of systems on the same data. Copyright restrictions and especially data privacy also interfere with the research progress [98]. Indeed, much more work is needed into evaluation measures to determine how valuable a text mining tool is for the actual user [98]. Some examples exist (e.g., for biocuration [99] and systematic reviews [100]) but limited work addresses deployment in a clinical context, which poses significant challenges. To find the right trade-off between the processing speed and the accuracy of a text mining algorithm is a challenging task, especially for online applications which must be responsive [101]. Furthermore, text mining tools should not only focus on English text documents, but should be able to process other languages too [102]. Recently, there have been some efforts to explore adaptation of predominantly English vocabulary resources and tools to other languages [103] and to address multilingual text processing in the health context [104] but there is significantly more to be done in this area, in particular to support integration of clinical information across language-specific data sets.

5.2 Specific Characteristics of Biomedical Texts

Clinical texts, in contrast to edited published articles, are in often not grammatically correct, use locally used abbreviations and have misspellings [11]. This poses huge challenges for transferring tools that have been developed for text analysis in the literature mining context to the clinical context. At a minimum, it requires appropriate normalization and spelling correction methods for text [105]. Even well-formed medical texts have different characteristics from general domain texts, and require tailored solutions. The BioLemmatizer [17] was developed specifically for handling inflectional morphology in biomedical texts; general English solutions do not have adequate coverage of the domain-specific vocabulary. Within clinical texts, there can be substantial variation in how 'regular' they are from a text processing perspective; emergency department triage notes written during a two-minute assessment of a patient will be substantially noisier than radiology reports or discharge summaries. More work is required to address adaptation of text processing tools to less well-formed texts. The biomedical literature also has certain characteristics which require specialised text processing strategies [106]. Enrichment and analysis of specific document parts such as tables gives access to a wealth of information, but cannot be handled using standard vector space or natural language processing strategies [107–110]. Similarly, it has recently been shown that substantial information in the biomedical literature actually resides in the *supplementary files* associated with publications rather than the main narrative text [111]; due to the varied and unpredictable nature of such files accessing this information will pose challenges to existing methods for information extraction.

5.3 Linguistic Issues and Semantic Enrichment

When extracting information from text documents one has to deal with several challenges arising from the complexities of natural language communication. One issue, which is particularly problematic for instance when trying to find relations among entities, is that language is very rich and one can express the same underlying concepts in many different ways. In addition, such relations expressed over multiple sentences which makes it even more difficult to find them and typically requires co-reference resolution. Ambiguity of words, phrases and entire sentences is one of the leading challenges in text mining that emerges because of the complexity of natural language itself. Some progress on ambiguity, especially term ambiguity, has been made; there exists a disambiguation module in the UMLS MetaMap concept recognition tool as well as other proposed approaches [112,113]. On the other hand, substantial challenges remain. For instance, although disambiguation of gene names in the biomedical literature (also known as gene normalization, in which mentions of genes are linked to a reference entity in a gene database) has been addressed in a recent challenge [114], the performance of the state-of-the-art systems has left room for improvement. Gene names are heavily overloaded, re-used across organisms and often confused with the function of the gene (e.g. "transcription growth factor"). An approach for entity disambiguation in the biomedical domain has been presented in [115]. Higher-level ambiguity such as processing of coordination structures and attachment of modifiers will require more linguistic methods and has been addressed in few systems [116,117] but remains a stumbling block. Co-reference resolution has been addressed in limited contexts [118–120] and also continues to pose a substantial challenge to integration of information from across a text.

6 Conclusion and Future Outlook

Biomedical text mining has benefited from substantial interest from the research community and from practical needs of the clinical domain, particularly in the last ten years. However, there remain substantial opportunities to progress the field further. One important direction is in the continued improvement of existing methods, supported through new creation of *gold standards* as benchmark sets [6]. Such resources are urgently needed to enable further improvement of methods, particularly in the clinical context where only a limited range of text types and tasks have been rigorously explored. Of particular importance are open data sets that can internationally be used [121], [106].

Another area of great opportunity is in methodological hot topics such as the graph-theoretical and topological text mining methods which are very promising approaches, yet not much studied [122]. Much potential for further research has the application of evolutionary algorithms [123], for text mining [124].

There are in addition a large number of opportunities to apply text mining to new problems and new text types. Text analysis of Web 2.0 and social media [125] is an emerging focus in biomedicine, for instance for detecting influenza-likes illnesses [126] or adverse drug events [127,128], [129].

In practice we need *integrated solutions* [130] of content analytics tools [131] into the clinical workplace. Integration of text mining with more general data mining is also a fruitful direction. In the clinical context drawing signals from both text, structured patient data (e.g. biometrics or laboratory results), and even biomedical images, will likely enable a more complete picture of a patient and his or her disease status for clinical decision support or outcome modeling. Such applications will require new strategies for multi-modal data integration and processing that incorporate text mining as a fundamental component. These tasks will result in solutions that will have substantial real-world impact and will highlight the importance of text mining for biomedicine.

Acknowledgments. We thank the Austrian IBM Watson Think Group for fruitful discussions and for the support of Gottfried Prohaska, Alberto Brabenetz and Michael Grosinger. We are grateful for comments from Michael Granitzer from the University of Passau, Germany. Part of the presented work was developed within the EEXCESS project funded by the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement number 600601.

References

1. Holzinger, A., Dehmer, M., Jurisica, I.: Knowledge discovery and interactive data mining in bioinformatics: State-of-the-art, future challenges and research directions. BMC Bioinformatics 15(suppl. 6), I1 (2014)
2. Holzinger, A.: Biomedical Informatics: Discovering Knowledge in Big Data. Springer, New York (2014)
3. Holzinger, A.: On Knowledge Discovery and Interactive Intelligent Visualization of Biomedical Data - Challenges in Human Computer Interaction and Biomedical Informatics, pp. 9–20. INSTICC, Rome (2012)

4. Holzinger, A., Stocker, C., Dehmer, M.: Big complex biomedical data: Towards a taxonomy of data. In: Springer Communications in Computer and Information Science. Springer, Heidelberg (in print, 2014)
5. Resnik, P., Niv, M., Nossal, M., Kapit, A., Toren, R.: Communication of clinically relevant information in electronic health records: a comparison between structured data and unrestricted physician language. In: CAC Proceedings of the Perspectives in Health Information Management (2008)
6. Kreuzthaler, M., Bloice, M., Faulstich, L., Simonic, K., Holzinger, A.: A comparison of different retrieval strategies working on medical free texts. Journal of Universal Computer Science 17(7), 1109–1133 (2011)
7. Holzinger, A., Geierhofer, R., Modritscher, F., Tatzl, R.: Semantic information in medical information systems: Utilization of text mining techniques to analyze medical diagnoses. Journal of Universal Computer Science 14(22), 3781–3795 (2008)
8. Witten, I., Frank, E., Hall, M.: Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, San Francisco (2011)
9. Verspoor, K., Cohen, K.: Natural language processing. In: Dubitzky, W., Wolkenhauer, O., Cho, K.H., Yokota, H. (eds.) Encyclopedia of Systems Biology, pp. 1495–1498. Springer, Heidelberg (2013)
10. Cohen, K.B., Demner-Fushman, D.: Biomedical Natural Language Processing. John Benjamins (2014)
11. Holzinger, A., Geierhofer, R., Errath, M.: Semantische Informationsextraktion in medizinischen Informationssystemen. Informatik Spektrum 30(2), 69–78 (2007)
12. Kumar, V., Tipney, H. (eds.): Biomedical Literature Mining. Methods in Molecular Biology, vol. 1159. Springer (2014)
13. Seifert, C., Sabol, V., Kienreich, W., Lex, E., Granitzer, M.: Visual analysis and knowledge discovery for text. In: Gkoulalas-Divanis, A., Labbi, A. (eds.) Large Scale Data Analytics, pp. 189–218. Springer (2014)
14. Manning, C.D., Raghavan, P., Schütze, H.: Introduction to Information Retrieval. Cambridge University Press, New York (2008)
15. W3C: HTML5 : a vocabulary and associated APIs for HTML and XHTML (2012)
16. Adobe Systems, I.: Pdf reference, 6th edn., version 1.23. (2006)
17. Liu, H., Christiansen, T., Baumgartner Jr., W.A., Verspoor, K.: BioLemmatizer: a lemmatization tool for morphological processing of biomedical text. Journal of Biomedical Semantics 3(3) (2012)
18. Porter, M.: An algorithm for suffix stripping. Program 14(3), 130–137 (1980)
19. Salton, G., Wong, A., Yang, C.: A vector space model for automatic indexing. Communications of the ACM 18(11), 620 (1975)
20. Boerjesson, E., Hofsten, C.: A vector model for perceived object rotation and translation in space. Psychological Research 38(2), 209–230 (1975)
21. Joachims, T.: Text categorization with suport vector machines: Learning with many relevant features. In: Nédellec, C., Rouveirol, C. (eds.) ECML 1998. LNCS, vol. 1398, pp. 137–142. Springer, Heidelberg (1998)
22. Crouch, C., Crouch, D., Nareddy, K.: Connectionist model for information retrieval based on the vector space model. International Journal of Expert Systems 7(2), 139–163 (1994)
23. Spärk Jones, K., Walker, S., Robertson, S.E.: A probabilistic model of information retrieval: development and comparative experiments. Inf. Process. Manage. 36(6) (2000)

24. Hliaoutakis, A., Varelas, G., Voutsakis, E., Petrakis, E., Milios, E.: Information Retrieval by Semantic Similarity. Intern. Journal on Semantic Web and Information Systems (IJSWIS) 3(3), 55–73 (2006); Special Issue of Multimedia Semantics
25. Hersh, W., Buckley, C., Leone, T.J., Hickam, D.: Ohsumed: An interactive retrieval evaluation and new large test collection for research. In: Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 1994, pp. 192–201. Springer-Verlag New York, Inc., New York (1994)
26. Müller, H., Michoux, N., Bandon, D., Geissbuhler, A.: A review of content-based image retrieval systems in medical applications - clinical benefits and future directions. International Journal of Medical Informatics 73(1), 1–23 (2003)
27. da Cunha, I., Fernández, S., Velázquez Morales, P., Vivaldi, J., SanJuan, E., Torres-Moreno, J.-M.: A new hybrid summarizer based on vector space model, statistical physics and linguistics. In: Gelbukh, A., Kuri Morales, Á.F. (eds.) MICAI 2007. LNCS (LNAI), vol. 4827, pp. 872–882. Springer, Heidelberg (2007)
28. Liu, G.: Semantic Vector Space Model: Implementation and Evaluation. Journal of the American Society for Information Science 48(5), 395–417 (1997)
29. Bellegarda, J.: Latent semantic mapping (information retrieval). IEEE Signal Processing Magazine 22(5), 70–80 (2005)
30. Landauer, T., Dumais, S.: A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychological Review 104(2), 211–240 (1997)
31. Landauer, T., Foltz, P., Laham, D.: An introduction to latent semantic analysis. Discourse Processes 25, 259–284 (1998)
32. Foltz, P., Kintsch, W., Landauer, T.: The measurement of textual coherence with latent semantic analysis. Discourse Processes 25, 285–308 (1998)
33. Kintsch, W.: The potential of latent semantic analysis for machine grading of clinical case summaries. Journal of Biomedical Informatics 35(1), 3–7 (2002)
34. Cohen, T., Blatter, B., Patel, V.: Simulating expert clinical comprehension: adapting latent semantic analysis to accurately extract clinical concepts from psychiatric narrative. Journal of Biomedical Informatics 41(6), 1070–1087 (2008)
35. Yeh, J.F., Wu, C.H., Chen, M.J.: Ontology-based speech act identification in a bilingual dialog system using partial pattern trees. J. Am. Soc. Inf. Sci. Technol. 59(5), 684–694 (2008)
36. Ginter, F., Suominen, H., Pyysalo, S., Salakoski, T.: Combining hidden markov models and latent semantic analysis for topic segmentation and labeling: Method and clinical application. I. J. Medical Informatics 78(12), 1–6 (2009)
37. Jonnalagadda, S., Cohen, T., Wu, S., Gonzalez, G.: Enhancing clinical concept extraction with distributional semantics. Journal of biomedical informatics 45(1), 129–140 (2012)
38. Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R.: Indexing by latent semantic analysis. Journal of the American Society for Information Science 41(6), 391–407 (1990)
39. Hofmann, T.: Probabilistic latent semantic indexing. In: Proceedings of the 22Nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 1999, pp. 50–57. ACM, New York (1999)
40. Papadimitriou, C., Raghavan, P., Tamaki, H., Vempala, S.: Latent semantic indexing: A probabilistic analysis. Journal of Computer and System Sciences 61(2), 217–235 (2000)
41. Hofmann, T.: Unsupervised Learning by Probabilistic Latent Semantic Analysis. Machine Learning 42, 177–196 (2001)

42. Xu, G., Zhang, Y., Zhou, X.: A web recommendation technique based on probabilistic latent semantic analysis. In: Ngu, A.H.H., Kitsuregawa, M., Neuhold, E.J., Chung, J.-Y., Sheng, Q.Z. (eds.) WISE 2005. LNCS, vol. 3806, pp. 15–28. Springer, Heidelberg (2005)
43. Si, L., Jin, R.: Adjusting mixture weights of gaussian mixture model via regularized probabilistic latent semantic analysis. In: Ho, T.-B., Cheung, D., Liu, H. (eds.) PAKDD 2005. LNCS (LNAI), vol. 3518, pp. 622–631. Springer, Heidelberg (2005)
44. Lin, C., Xue, G., Zeng, H., Yu, Y.: Using Probabilistic Latent Semantic Analysis for Personalized Web Search. In: Zhang, Y., Tanaka, K., Yu, J.X., Wang, S., Li, M. (eds.) APWeb 2005. LNCS, vol. 3399, pp. 707–717. Springer, Heidelberg (2005)
45. Kim, Y.S., Oh, J.S., Lee, J.Y., Chang, J.H.: An intelligent grading system for descriptive examination papers based on probabilistic latent semantic analysis. In: Webb, G.I., Yu, X. (eds.) AI 2004. LNCS (LNAI), vol. 3339, pp. 1141–1146. Springer, Heidelberg (2004)
46. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B 39, 1–38 (1977)
47. Dobrokhotov, P.B., Goutte, C., Veuthey, A.L., Gaussier, R.: Assisting medical annotation in swiss-prot using statistical classifiers. I. J. Medical Informatics 74(2-4), 317–324 (2005)
48. Srinivas, K., Rao, G., Govardhan, A.: Survey on prediction of heart morbidity using data mining techniques. International Journal of Data Mining & ... 1(3), 14–34 (2011)
49. Lu, Y., Zhang, P., Deng, S.: Exploring Health-Related Topics in Online Health Community Using Cluster Analysis. In: 2013 46th Hawaii International Conference on System Sciences, pp. 802–811 (January 2013)
50. Masseroli, M., Chicco, D., Pinoli, P.: Probabilistic latent semantic analysis for prediction of gene ontology annotations. In: The 2012 International Joint Conference on Neural Networks (IJCNN), pp. 1–8 (2012)
51. Koehler, R.: Aspects of Automatic Text Analysis. Springer (2007)
52. Blei, D., Ng, A., Jordan, M.: Latent dirichlet allocation. The Journal of Machine Learning Research 3, 993–1022 (2003)
53. Kakkonen, T., Myller, N., Sutinen, E.: Applying latent Dirichlet allocation to automatic essay grading. In: Salakoski, T., Ginter, F., Pyysalo, S., Pahikkala, T. (eds.) FinTAL 2006. LNCS (LNAI), vol. 4139, pp. 110–120. Springer, Heidelberg (2006)
54. Xing, D., Girolami, M.: Employing latent dirichlet allocation for fraud detection in telecommunications. Pattern Recognition Letters 28(13), 1727–1734 (2007)
55. Girolami, M., Kaban, A.: Sequential activity profiling: Latent Dirichlet allocation of Markov chains. Data Mining and Knowledge Discovery 10(3), 175–196 (2005)
56. Minka, T., Lafferty, J.: Expectation-propagation for the generative aspect model. In: Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence, UAI 2002, pp. 352–359. Morgan Kaufmann Publishers Inc., San Francisco (2002)
57. Griffiths, T.L., Steyvers, M.: Finding scientific topics. Proceedings of the National Academy of Sciences 101(suppl. 1), 5228–5235 (2004)
58. Asou, T., Eguchi, K.: Predicting protein-protein relationships from literature using collapsed variational latent dirichlet allocation. In: Proceedings of the 2nd International Workshop on Data and Text Mining in Bioinformatics, DTMBIO 2008, pp. 77–80. ACM, New York (2008)

59. Arnold, C.W., El-Saden, S.M., Bui, A.A.T., Taira, R.: Clinical case-based retrieval using latent topic analysis. In: AMIA Annu. Symp. Proc., vol. 2010, pp. 26–30 (2010)
60. Arnold, C., Speier, W.: A topic model of clinical reports. In: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2012, pp. 1031–1032. ACM, New York (2012)
61. Yao, L., Riedel, S., McCallum, A.: Unsupervised relation discovery with sense disambiguation. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers, ACL 2012, vol. 1, pp. 712–720. Association for Computational Linguistics, Stroudsburg (2012)
62. Dawson, J., Kendziorski, C.: Survival-supervised latent Dirichlet allocation models for genomic analysis of time-to-event outcomes. arXiv preprint arXiv:1202.5999, 1–21 (2012)
63. Hripcsak, G., Albers, D.J.: Next-generation phenotyping of electronic health records. JAMIA 20(1), 117–121 (2013)
64. Blei, D.M., Griffiths, T.L., Jordan, M.I.: The nested chinese restaurant process and bayesian nonparametric inference of topic hierarchies. J. ACM 57(2), 7:1–7:30 (2010)
65. Pitman, J.: Combinatorial stochastic processes. Springer Lecture Notes in Mathematics. Springer (2002); Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour (2002)
66. Saria, S., Koller, D., Penn, A.: Discovering shared and individual latent structure in multiple time series. arXiv preprint arXiv:1008 (d), 1–9 (2028)
67. Bartlett, N., Wood, F., Perotte, A.: Hierarchically Supervised Latent Dirichlet Allocation. In: NIPS, pp. 1–9 (2011)
68. Hotelling, H.: Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology 24(6), 417–441 (1933)
69. Pearson, K.: LIII. On lines and planes of closest fit to systems of points in space. Philosophical Magazine Series 6 2(11), 559–572 (1901)
70. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., Reich, D.: Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38(8), 904–909 (2006)
71. Robertson, M.M., Althoff, R.R., Hafez, A., Pauls, D.L.: Principal components analysis of a large cohort with Tourette syndrome. The British Journal of Psychiatry: the Journal of Mental Science 193(1), 31–36 (2008)
72. Himmel, W., Reincke, U., Michelmann, H.W.: Text mining and natural language processing approaches for automatic categorization of lay requests to web-based expert forums. Journal of Medical Internet Research 11(3), e25 (2009)
73. Oprea, T., Nielsen, S., Ursu, O.: Associating Drugs, Targets and Clinical Outcomes into an Integrated Network Affords a New Platform for Computer Aided Drug Repurposing. Molecular Informatics 30, 100–111 (2011)
74. Schölkopf, B., Smola, A., Müller, K.R.: Kernel principal component analysis. In: Gerstner, W., Hasler, M., Germond, A., Nicoud, J.-D. (eds.) ICANN 1997. LNCS, vol. 1327, pp. 583–588. Springer, Heidelberg (1997)
75. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. 20(3), 273–297 (1995)
76. Ben-Hur, A., Weston, J.: A user's guide to support vector machines. In: Carugo, O., Eisenhaber, F. (eds.) Data Mining Techniques for the Life Sciences. Methods in Molecular Biology, vol. 609, pp. 223–239. Humana Press (2010)
77. Crammer, K., Singer, Y.: On the algorithmic implementation of multiclass kernel-based vector machines. J. Mach. Learn. Res. 2, 265–292 (2002)

78. Guyon, I., Weston, J., Barnhill, S., Vapnik, V.: Gene selection for cancer classification using support vector machines. Mach. Learn. 46(1-3), 389–422 (2002)
79. Ghanem, M., Guo, Y., Lodhi, H., Zhang, Y.: Automatic scientific text classification using local patterns: Kdd cup 2002 (task 1). SIGKDD Explorations 4(2), 95–96 (2002)
80. Donaldson, I.M., Martin, J.D., de Bruijn, B., Wolting, C., Lay, V., Tuekam, B., Zhang, S., Baskin, B., Bader, G.D., Michalickova, K., Pawson, T., Hogue, C.W.V.: Prebind and textomy - mining the biomedical literature for protein-protein interactions using a support vector machine. BMC Bioinformatics 4, 11 (2003)
81. Eskin, E., Agichtein, E.: Combining text mining and sequence analysis to discover protein functional regions. In: Altman, R.B., Dunker, A.K., Hunter, L., Jung, T.A., Klein, T.E. (eds.) Pacific Symposium on Biocomputing, pp. 288–299. World Scientific (2004)
82. Joshi, M., Pedersen, T., Maclin, R.: A comparative study of support vector machines applied to the supervised word sense disambiguation problem in the medical domain. In: Prasad, B. (ed.) IICAI, pp. 3449–3468 (2005)
83. Uzuner, Z., Bodnari, A., Shen, S., Forbush, T., Pestian, J., South, B.R.: Evaluating the state of the art in coreference resolution for electronic medical records. JAMIA 19(5), 786–791 (2012)
84. Domingos, P., Pazzani, M.: On the optimality of the simple bayesian classifier under zero-one loss. Mach. Learn. 29(2-3), 103–130 (1997)
85. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers Inc., San Francisco (1993)
86. Kim, J.D., Pyysalo, S.: Bionlp shared task. In: Dubitzky, W., Wolkenhauer, O., Cho, K.H., Yokota, H. (eds.) Encyclopedia of Systems Biology, pp. 138–141. Springer, New York (2013)
87. Arighi, C., Lu, Z., Krallinger, M., Cohen, K., Wilbur, W., Valencia, A., Hirschman, L., Wu, C.: Overview of the biocreative iii workshop. BMC Bioinformatics 12(suppl. 8), S1 (2011)
88. Hersh, W., Voorhees, E.: Trec genomics special issue overview. Information Retrieval 12(1), 1–15 (2009)
89. Kim, J.D., Ohta, T., Tateisi, Y., Tsujii, J.: Genia corpus: a semantically annotated corpus for bio-textmining. Bioinformatics 19(suppl. 1), i180–i182 (2003)
90. Bada, M., Eckert, M., Evans, D., Garcia, K., Shipley, K., Sitnikov, D., Baumgartner, W., Cohen, K., Verspoor, K., Blake, J., Hunter, L.: Concept annotation in the CRAFT corpus. BMC Bioinformatics 13(161) (2012)
91. Verspoor, K., Cohen, K., Lanfranchi, A., Warner, C., Johnson, H., Roeder, C., Choi, J., Funk, C., Malenkiy, Y., Eckert, M., Xue, N., Baumgartner, W., Bada, M., Palmer, M., Hunter, L.: A corpus of full-text journal articles is a robust evaluation tool for revealing differences in performance of biomedical natural language processing tools. BMC Bioinformatics 13, 207 (2012)
92. Klinger, R., Kolik, C., Fluck, J., Hofmann-Apitius, M., Friedrich, C.M.: Detection of iupac and iupac-like chemical names. Bioinformatics 24(13), i268–i276 (2008)
93. Verspoor, K., Jimeno Yepes, A., Cavedon, L., McIntosh, T., Herten-Crabb, A., Thomas, Z., Plazzer, J.P.: Annotating the biomedical literature for the human variome. Database 2013 (2013)
94. Voorhees, E., Tong, R.: Overview of the trec 2011 medical records track. In: Proceedings of the Text Retrieval Conference (2011)
95. Uzuner, O.: Second i2b2 workshop on natural language processing challenges for clinical records. In: Proceedings of the American Medical Informatics Association Annual Symposium, pp. 1252–1253 (2008)

96. Sun, W., Rumshisky, A., Uzuner, O.: Evaluating temporal relations in clinical text: 2012 i2b2 challenge. Journal of the American Medical Informatics Association 20(5), 806–813 (2013)
97. Suominen, H., et al.: Overview of the share/clef ehealth evaluation lab 2013. In: Forner, P., Müller, H., Paredes, R., Rosso, P., Stein, B. (eds.) CLEF 2013. LNCS, vol. 8138, pp. 212–231. Springer, Heidelberg (2013)
98. Cohen, A.M., Hersh, W.R.: A survey of current work in biomedical text mining. Briefings in Bioinformatics 6(1), 57–71 (2005)
99. Hirschman, L., Burns, G.A.P.C., Krallinger, M., Arighi, C., Cohen, K.B., Valencia, A., Wu, C.H., Chatr-Aryamontri, A., Dowell, K.G., Huala, E., Loureno, A., Nash, R., Veuthey, A.L., Wiegers, T., Winter, A.G.: Text mining for the biocuration workflow. Database 2012 (2012)
100. Ananiadou, S., Rea, B., Okazaki, N., Procter, R., Thomas, J.: Supporting systematic reviews using text mining. Social Science Computer Review 27(4), 509–523 (2009)
101. Dai, H.J., Chang, Y.C., Tsai, R.T.H., Hsu, W.L.: New challenges for biological text-mining in the next decade. J. Comput. Sci. Technol. 25(1), 169–179 (2009)
102. Tan, A.H.: Text mining: The state of the art and the challenges. In: Proceedings of the Pacific Asia Conf on Knowledge Discovery and Data Mining PAKDD 1999, Workshop on Knowledge Discovery from Advanced Databases, KDAD 1999, pp. 65–70 (1999)
103. Carrero, F., Cortizo, J., Gomez, J.: Testing concept indexing in crosslingual medical text classification. In: Third International Conference on Digital Information Management, ICDIM 2008, pp. 512–519 (November 2008)
104. Allvin, H., Carlsson, E., Dalianis, H., Danielsson-Ojala, R., Daudaravičius, V., Hassel, M., Kokkinakis, D., Lundgren-Laine, H., Nilsson, G., Nytrø, O., Salanterä, S., Skeppstedt, M., Suominen, H., Velupillai, S.: Characteristics and analysis of finnish and swedish clinical intensive care nursing narratives. In: Proceedings of the NAACL HLT 2010 Second Louhi Workshop on Text and Data Mining of Health Documents, Louhi 2010, pp. 53–60. Association for Computational Linguistics, Stroudsburg (2010)
105. Patrick, J., Sabbagh, M., Jain, S., Zheng, H.: Spelling correction in clinical notes with emphasis on first suggestion accuracy. In: 2nd Workshop on Building and Evaluating Resources for Biomedical Text Mining, pp. 2–8 (2010)
106. Holzinger, A., Yildirim, P., Geier, M., Simonic, K.M.: Quality-based knowledge discovery from medical text on the web. In: Pasi, G., Bordogna, G., Jain, L.C. (eds.) Quality Issues in the Management of Web Information, Intelligent Systems Reference Library. ISRL, vol. 50, pp. 145–158. Springer, Heidelberg (2013)
107. Wong, W., Martinez, D., Cavedon, L.: Extraction of named entities from tables in gene mutation literature. In: BioNLP 2009, p. 46 (2009)
108. Limaye, G., Sarawagi, S., Chakrabarti, S.: Annotating and searching web tables using entities, types and relationships. In: Proc. VLDB Endow., vol. 3(1-2), pp. 1338–1347 (September 2010)
109. Quercini, G., Reynaud, C.: Entity discovery and annotation in tables. In: Proceedings of the 16th International Conference on Extending Database Technology, EDBT 2013, pp. 693–704. ACM, New York (2013)
110. Zwicklbauer, S., Einsiedler, C., Granitzer, M., Seifert, C.: Towards disambiguating web tables. In: International Semantic Web Conference (Posters & Demos), pp. 205–208 (2013)

111. Jimeno Yepes, A., Verspoor, K.: Literature mining of genetic variants for curation: Quantifying the importance of supplementary material. Database: The Journal of Biological Databases and Curation 2013 (2013)
112. Liu, H., Johnson, S.B., Friedman, C.: Automatic Resolution of Ambiguous Terms Based on Machine Learning and Conceptual Relations in the UMLS. Journal of the American Medical Informatics Association 9(6), 621–636 (2002)
113. Aronson, A.R., Lang, F.M.: An overview of metamap: historical perspective and recent advances. Journal of the American Medical Informatics Association 17(3), 229–236 (2010)
114. Lu, Z., Kao, H.Y., Wei, C.H., Huang, M., Liu, J., Kuo, C.J., Hsu, C.N., Tsai, R., Dai, H.J., Okazaki, N., Cho, H.C., Gerner, M., Solt, I., Agarwal, S., Liu, F., Vishnyakova, D., Ruch, P., Romacker, M., Rinaldi, F., Bhattacharya, S., Srinivasan, P., Liu, H., Torii, M., Matos, S., Campos, D., Verspoor, K., Livingston, K., Wilbur, W.: The gene normalization task in biocreative iii. BMC Bioinformatics 12(suppl. 8), S2 (2011)
115. Zwicklbauer, S., Seifert, C., Granitzer, M.: Do we need entity-centric knowledge bases for entity disambiguation? In: Proceedings of the 13th International Conference on Knowledge Management and Knowledge Technologies, I-Know (2013)
116. Ogren, P.V.: Improving syntactic coordination resolution using language modeling. In: Proceedings of the NAACL HLT 2010 Student Research Workshop, HLT-SRWS 2010, pp. 1–6. Association for Computational Linguistics, Stroudsburg (2010)
117. Chae, J., Jung, Y., Lee, T., Jung, S., Huh, C., Kim, G., Kim, H., Oh, H.: Identifying non-elliptical entity mentions in a coordinated {NP} with ellipses. Journal of Biomedical Informatics 47, 139–152 (2014)
118. Gasperin, C., Briscoe, T.: Statistical anaphora resolution in biomedical texts. In: Proceedings of the 22nd International Conference on Computational Linguistics, COLING 2008, vol. 1, pp. 257–264. Association for Computational Linguistics, Stroudsburg (2008)
119. Jonnalagadda, S.R., Li, D., Sohn, S., Wu, S.T.I., Wagholikar, K., Torii, M., Liu, H.: Coreference analysis in clinical notes: a multi-pass sieve with alternate anaphora resolution modules. Journal of the American Medical Informatics Association 19(5), 867–874 (2012)
120. Kim, J.D., Nguyen, N., Wang, Y., Tsujii, J., Takagi, T., Yonezawa, A.: The genia event and protein coreference tasks of the bionlp shared task 2011. BMC Bioinformatics 13(suppl. 11), S1 (2012)
121. Yildirim, P., Ekmekci, I.O., Holzinger, A.: On knowledge discovery in open medical data on the example of the fda drug adverse event reporting system for alendronate (fosamax). In: Holzinger, A., Pasi, G. (eds.) HCI-KDD 2013. LNCS, vol. 7947, pp. 195–206. Springer, Heidelberg (2013)
122. Holzinger, A.: On topological data mining. In: Holzinger, A., Jurisica, I. (eds.) Knowledge Discovery and Data Mining. LNCS, vol. 8401, pp. 333–358. Springer, Heidelberg (2014)
123. Holzinger, K., Palade, V., Rabadan, R., Holzinger, A.: Darwin or lamarck? future challenges in evolutionary algorithms for knowledge discovery and data mining. In: Holzinger, A., Jurisica, I. (eds.) Knowledge Discovery and Data Mining. LNCS, vol. 8401, pp. 35–56. Springer, Heidelberg (2014)
124. Mukherjee, I., Al-Fayoumi, M., Mahanti, P., Jha, R., Al-Bidewi, I.: Content analysis based on text mining using genetic algorithm. In: 2nd International Conference on Computer Technology and Development (ICCTD), pp. 432–436. IEEE (2010)

125. Petz, G., Karpowicz, M., Fürschuß, H., Auinger, A., Stříteský, V., Holzinger, A.: Opinion mining on the web 2.0 – characteristics of user generated content and their impacts. In: Holzinger, A., Pasi, G. (eds.) HCI-KDD 2013. LNCS, vol. 7947, pp. 35–46. Springer, Heidelberg (2013)
126. Corley, C.D., Cook, D.J., Mikler, A.R., Singh, K.P.: Text and structural data mining of influenza mentions in Web and social media. International Journal of Environmental Research and Public Health 7(2), 596–615 (2010)
127. White, R.W., Tatonetti, N.P., Shah, N.H., Altman, R.B., Horvitz, E.: Web-scale pharmacovigilance: listening to signals from the crowd. Journal of the American Medical Informatics Association (2013)
128. Wu, H., Fang, H., Stanhope, S.J.: Exploiting online discussions to discover unrecognized drug side effects. Methods of Information in Medicine 52(2), 152–159 (2013)
129. Yildirim, P., Majnaric, L., Ekmekci, O., Holzinger, A.: Knowledge discovery of drug data on the example of adverse reaction prediction. BMC Bioinformatics 15(suppl. 6), S7 (2014)
130. Holzinger, A., Jurisica, I.: Knowledge discovery and data mining in biomedical informatics: The future is in integrative, interactive machine learning solutions. In: Holzinger, A., Jurisica, I. (eds.) Knowledge Discovery and Data Mining. LNCS, vol. 8401, pp. 1–18. Springer, Heidelberg (2014)
131. Holzinger, A., Stocker, C., Ofner, B., Prohaska, G., Brabenetz, A., Hofmann-Wellenhof, R.: Combining hci, natural language processing, and knowledge discovery - potential of ibm content analytics as an assistive technology in the biomedical domain. In: Holzinger, A., Pasi, G. (eds.) HCI-KDD 2013. LNCS, vol. 7947, pp. 13–24. Springer, Heidelberg (2013)