Detection of Privacy Disclosure in the Medical Domain: A Survey

Bianca Buff, Joschka Kersting and Michaela Geierhos
Semantic Information Processing Group, Paderborn University, Warburger Str. 100, Paderborn, Germany

Keywords: Identity Disclosure, Privacy Protection, Physician Review Website, De-anonymization, Medical Domain.

Abstract: When it comes to increased digitization in the health care domain, privacy is a relevant topic nowadays. This relates to patient data, electronic health records or physician reviews published online, for instance. There exist different approaches to the protection of individuals privacy, which focus on the anonymization and masking of personal information subsequent to their mining. In the medical domain in particular, measures to protect the privacy of patients are of high importance due to the amount of sensitive data that is involved (e.g. age, gender, illnesses, medication). While privacy breaches in structured data can be detected more easily, disclosure in written texts is more difficult to find automatically due to the unstructured nature of natural language. Therefore, we take a detailed look at existing research on areas related to privacy protection. Likewise, we review approaches to the automatic detection of privacy disclosure in different types of medical data. We provide a survey of several studies concerned with privacy breaches in the medical domain with a focus on Physician Review Websites (PRWs). Finally, we briefly develop implications and directions for further research.

1 INTRODUCTION

A website where users can create profiles and choose pseudonyms, leads them to assume a high degree of anonymity. If users compose posts under a pseudonym or even altogether anonymized without publishing any name, they feel safe to disclose private information. However, users are often not aware of the fact that even a limited amount of personal information can lead to identification (Bäumer et al., 2017; Yang et al., 2012; Kersting et al., 2019). Consequently, this gap between perceived and actual anonymity is problematic when users disclose information unintentionally, accidentally and/or without taking note of it. This is more likely to happen in an online environment such as a Physician Review Websites (PRW) than in a less anonymized setting where they probably would not have revealed this private information (Bäumer et al., 2017).

Privacy is a hot topic that raises concerns and controversies. Hence, measures and laws to protect individuals' privacy are constantly created and refined. This applies not only to health data intended to be investigated, published or shared, but also to other data published on the Web, e.g. on social networking sites or forums. The medical domain specifically deals with a high amount of sensitive data. Consequently, privacy protection is an important subject in this sector, because privacy disclosure can have many negative impacts on the person concerned (Cofone, 2017; Bäumer et al., 2017).

The aim of this paper is to provide an overview of existing research in the field of privacy disclosure and privacy protection in the medical domain. Several studies have focused on these topics (Mendes and Vilela, 2017; Dankar and El Emam, 2012; Bäumer et al., 2017; Li et al., 2010). The majority of research focuses on structured data that is often created by professionals, such as physicians or in contexts of medical studies (El Emam et al., 2009; Gal et al., 2008; Kim et al., 2014). However, there are large amounts of textual, unstructured data such as review texts on PRWs which so far cannot be anonymized easily (Bäumer et al., 2017). Privacy protection is a challenge here.

Existing privacy protection measurements such as k-anonymity (Samarati and Sweeney, 1998) focus on structured data. In contrast, natural language contains a "large number of unstructured (not predefined) sensitive attributes" (Li and Qin, 2017), hence those models are not applicable on this type of data. In the medical domain specifically, besides sensitive fea-

Buff, B., Kersting, J. and Geierhos, M.
Detection of Privacy Disclosure in the Medical Domain: A Survey.
DOI: 10.5220/0009347506300637
In *Proceedings of the 9th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2020)*, pages 630-637
ISBN: 978-989-758-397-1; ISSN: 2184-4313

tures such as names of persons and places, information about patients symptoms, diagnoses and treatments is disclosed. The biggest obstacle of privacy protection in natural language data is therefore the detection of sensitive information in the free text in the first place. Only subsequently can the disclosure of attributes be prevented.

The corresponding paper is structured as follows: First, the background and terminological information are provided on the relevant concepts. Then, challenges that arise in the research field of privacy protection are discussed before some computational approaches and applications concerning the detection of privacy disclosure and anonymization are presented. Finally, further directions and perspectives are summarized in the discussion and conclusion.

2 STATE-OF-THE-ART

The concept of *privacy* is not easy to grasp and not defined comprehensively and universally in the literature. Instead, various definitions of privacy co-exist, depending on the data concerned and the specific use case. Mendes and Vilela (2017) provide an overview and conclude that "the main idea of information privacy is to have control over the collection and handling of one's personal data".

When sensitive information is revealed, this can potentially be harmful for the individuals involved. A differentiation can be made between identity and attribute disclosure. The first means that due to information leakage, a certain individual can be identified out of a set of people. Attribute disclosure signifies that for a certain individual (known or unknown), the value of a certain attribute can be inferred (Duncan and Lambert, 1989). An example would be when the data shows that every 30-year-old male patient who went to a specific hospital was presented with the diagnosis of a certain illness. Then it can be concluded that a man who is 30 years old and was treated in this hospital must have had this diagnosis. The conclusion can be drawn no matter if the individual's concrete identity is known or not (El Emam, 2011).

In the literature, there is the differentiation between explicit identifiers, quasi-identifiers and sensitive attributes (Mendes and Vilela, 2017). Explicit identifiers directly allow the identification of an individual, i.e. names and medical health numbers fall in this category. Quasi-identifiers (e.g. age, gender) are non-sensitive insofar that when considered separately, they do not link directly to a specific individual. Sensitive attributes are confidential and specific to individuals, for instance a certain disease of a certain patient. Taken separately, quasi-identifiers and sensitive attributes do not directly lead to an individual's identification. Nevertheless, linkage attacks are still possible, i.e. identification attacks exploiting the combination of attributes or additional background knowledge (Fung et al., 2010). This is especially threatening when the data set contains many variables and unique combinations of attributes occur, which facilitates identification (Martnez et al., 2012). De-identification of a data set always includes the masking of all explicit identifiers (El Emam, 2006) and often the additional anonymization of quasi-identifiers and sensitive attributes (Fung et al., 2010; Snchez et al., 2013).

It is not clear-cut what exactly counts as a sensitive attribute. Some systems that focus on privacy in a broader sense and that apply machine learning classification algorithms treat all named entities (e.g. names, locations) as sensitive (Snchez et al., 2013). Regarding official regulations for privacy protection, governments play an important role in defining sensitivity. In the United States of America for instance, the *Health Insurance Portability and Accountability Act* (HIPAA) describes the regulations and limitations concerning private information (HIPAA, 1996). This act defines 18 pieces of *Protected Health Information* (PHI) that have to be masked in any medical document prior their publication in order to preserve the patients' identities. Examples for PHI identifiers are names, email addresses, dates and social security or medical health numbers.

With regard to means to protect privacy, the aim of so-called *privacy-preserving data mining* (PPDM) techniques is to mine data while ensuring that no critical amount of sensitive information is disclosed. In this context, the term *utility* describes the data quality remaining after the application of the privacy protection measure(s). This includes a natural trade-off between the information loss and the level of privacy (Mendes and Vilela, 2017). Ideally, the measures taken to protect privacy should minimize the information loss while maximizing the data utility (Martnez et al., 2011).

To protect the individuals' identities in a data set, several operations on the data can be applied (Fung et al., 2010): Such as suppression (i.e. the entire removal) of sensitive attributes or generalization, meaning that exact numerical values are generalized to a broader interval of values. Categorical data can be generalized to a broader term, e.g. the specific profession *mechanical engineer* to the hyperonym *engineer*. In addition, perturbation can be applied, meaning the replacement of the original data. More concretely, the values can be randomized, swapped or supplemented

with additional synthetically generated data. This is not an exhaustive list of possible operations; a more detailed overview is provided by Fung et al. (2010).

Implementations for data anonymization, called *privacy models* or *privacy metrics*, aim at protecting identity, typically by making use of the masking operations described above. Popular models are for instance *k*-anonymity (Samarati and Sweeney, 1998) and *l*-diversity (Machanavajjhala et al., 2006). These models stem from the database domain and "compute the level of privacy depending only on properties of the data" (Wagner and Eckhoff, 2018). Consequently, anonymization can be applied to all individuals in the database, for instance by generalizing a certain sensitive attribute so that individuals cannot be identified.

In the medical domain, PRWs are receiving increasing interest by patients and the research community (Emmert et al., 2012). These websites are a format for Internet users to share experiences about the perceived quality of received medical care. Online on a website such as the German PRW *jameda.de*[1], users can give both a quantitative and a qualitative rating of their treatment. More concretely, they not only give the physician grades on various pre-defined dimensions (e.g. friendliness, treatment, practice equipment), but also report qualitatively on the provided health care in the form of (written) free text.

Research concerning PRWs has focused on aspects such as the choice-making process of patients: Several studies (Emmert et al., 2009; McLennan et al., 2017; Okike et al., 2016) come to the conclusion that PRWs should not be used uniquely as a measure to finding a good and suitable physician. Reasons for that include that there is not necessarily a correlation between the rating and the actual quality (Okike et al., 2016) and that only a minority of patients writes reviews on PRWs, which limits the representativeness (McLennan et al., 2017). Furthermore, some studies investigated the content available on different review websites and studied in detail and qualitatively what patients say about their physicians (Emmert et al., 2012, 2014). Overall, the results indicate that patients tend to assess the physicians positively in their reviews (Emmert et al., 2013; Ellimoottil et al., 2013; Kadry et al., 2011). Notions of trust and relationships between physicians and patients have been focused on as well (Kersting et al., 2019). Increasingly, PRWs are investigated under the aspect of sentiment analysis, aiming at identifying aspects and sentiments in the free text automatically (Brody and Elhadad, 2010).

[1] Available at https://www.jameda.de.

3 CHALLENGES

In the following, we name several challenges regarding the concept of privacy, sensitive attributes and natural language data. A fundamental challenge is the definition of privacy and sensitive attributes according to the concrete data (Shah and Gulati, 2016). Depending on the respective domain, privacy concepts have to focus on different aspects. For instance, a location-based service uses information from the GPS module of the user's smart mobile device. This presents a privacy threat insofar that the constant collection of GPS data can provide insights into details such as the user's home and office address (Han et al., 2018). Consequently, an approach to protect user's privacy in a location-based service has to tackle different challenges than an approach concerned with privacy protection in the medical domain.

With the enhanced and more sophisticated data mining techniques, data collection has become easier and more large-scale. Yang et al. (2012) show that even a limited amount of (seed) information accessible on social networks can be enough to identify the majority of users. Li et al. (2010) come to a similar conclusion: It is easily possible to utilize various sources (e.g. different social networks, commercial search engines) to link the sensible pieces of information to the corresponding identities.

In the medical domain, even though regulations such as the HIPAA give concrete definitions of identifiers that always have to be protected, such as names and medical health numbers, the masking of just these attributes is often not enough. Snchez et al. (2013) found out that there exist semantic correlations between sanitized (i.e. masked in some way) and not sanitized attributes that can potentially be exploited and lead to privacy threats. In addition, Dankar and El Emam (2012) report of correlations between different data fields, such as drugs and diagnoses. Here, inconsistent anonymization can lead to distortions and contradictions. Then, the utility of and the trust in the data is decreased. Consequently, more research needs to be done to investigate not only the potential threats posed by PHI identifiers. Attributes going beyond them have to be taken into consideration as well, just like the underlying correlations existing between several attributes. It is still a matter of research to define the critical amount of revealed information that threatens privacy.

Concerning other types of data, a challenge for privacy protection is the nature of unstructured natural language data. The automatic detection of sensitive attributes in unstructured data, such as health reports, is difficult (Bäumer et al., 2017; Ganu et al.,

Table 1: Selection of studies focusing on de-identification in structured and unstructured medical data.

Authors	Data	Details of the method
Structured Data:		
Ayala-Rivera et al. (2014)	health data	anonymization using differential privacy
Gal et al. (2008)	health data	extended model of *k*-anonymity and *l*-diversity
El Emam et al. (2009)	health data	extended model of *k*-anonymity: Optimal Lattice Anonymization
Kim et al. (2014)	health data streams	delay-free online anonymization & late validation
Unstructured Data		
Snchez et al. (2014)	electronic healthcare records	automatic sanitization with usage of external knowledge bases (e.g. WordNet, SNOMED-CT)
Meystre et al. (2010)	electronic healthcare records	survey of dictionary-based and machine learning-based techniques
Li and Qin (2017)	medical text records	clustering & value enumeration
Luo et al. (2016)	medical text records	double-reading/entry system for creation of a database and modularization for anonymization
Meystre et al. (2014)	clinical notes	comparison of different text de-identification systems
Dernoncourt et al. (2017)	patient notes	bi-directional LSTM neural network
Gardner and Xiong (2008)	pathology reports	conditional random field for attribute detection and subsequent *k*-anonymization
Bäumer et al. (2017)	physician review websites	named entity recognition, lexical patterns, string similarity algorithms

2012). The PHI identifiers defined in the HIPAA generally follow a regular structure, e.g. ZIP codes or emails, or stem from a finite set of options, e.g. locations (Snchez et al., 2014). In contrast, attributes such as diseases or medication can be expressed in various ways in natural language. The missing consistent surface structure complicates pattern matching and machine learning approaches that are more easily applicable to structured PHI identifiers (Snchez et al., 2014). The following review from a PRW contains several privacy breaches: *I live down the street of Dr. Mayers practice and go there since I was a little child. I am 39 years old and since I gave birth to a little daughter five years ago, we visit Dr. Mayer together. He also helped me with my hemorrhoids last year, even though he practices in an utterly different field.* In this review, a woman honestly talks about the good performance of her health care provider (HCP). On the one hand, the details she adds about herself and her medical history make the review more credible and valuable. On the other hand, however, the sensitive attributes that she reveals can lead to identification by people she knows or even by strangers having additional background knowledge. Moreover, the HCP himself would probably be able to identify her, since HCPs always have a lot of (sensitive) information about their patients at their disposal.

The high variance in natural language data complicates the automatic detection of privacy breaches. Examples are words like *mother* or *father*: The semantics inherent to these words disclose the gender of the person (female or male) and their family status (at least one child) (Bäumer et al., 2017). In natural language, it is not clear-cut and well-defined what constitutes as a sensitive attribute that discloses too much information about a certain individual. Another challenge inherent to user-generated content is the following: Online reviews about the medical services that patients received typically incorporate a large amount of grammatical and typographical errors, in addition to idiosyncrasies (e.g. in the usage of abbreviations) (Bäumer et al., 2017). Even in the case of medical reports about patients written by doctors, where one might expect a shared standard and higher quality, there tend to be similar mistakes, peculiarities and sublanguage characteristics, due to the time pressure and other constraints (Pestian et al., 2007). These inconsistencies complicate the automatic detection of the privacy disclosures.

4 APPROACHES AND APPLICATIONS

This section introduces several approaches to and applications of privacy protection and de-anonymization in the medical domain. Table 1 provides an overview of the presented studies.

4.1 Structured Data

Privacy protection in structured medical data relies for instance on the privacy-preserving models *k*-anonymity and *l*-diversity (Ayala-Rivera et al., 2014; Dankar and El Emam, 2012). There exist several improvements of these models, such as the one by Gal et al. (2008). The authors extend the models to structured data that has multiple sensitive attributes, and not just one. This is sensible since data about patients is generally so high-dimensional that it includes more than one sensitive attribute. Another study introduced an extension of the *k*-anonymity algorithm that is applicable to health data sets, the *Optimal Lattice Anonymization* algorithm (El Emam et al., 2009). This improvement of *k*-anonymity decreases the information loss and produces globally optimal de-identification results.

Kim et al. (2014) focus on the online anonymization of health data streams. The individual inputs are anonymized immediately with counterfeit values, so there is no delay in the continuous transmission of data. This approach differs from previous ones where a certain amount of data has to be extracted before privacy-preserving measures can be applied on the respective batches, leading to a significant time delay. Additionally, the authors ensure a high data utility by applying late validation to limit the amount of created counterfeit values.

4.2 Unstructured Data

Privacy protection in unstructured data typically focuses on a certain clinical document type: e.g. on medical patient records, discharge summaries or clinical free text (Meystre et al., 2014). Considering aspects such as the uniqueness of answers to open questions possibly leading to identification, textual values can be regarded as quasi-identifiers that have to be anonymized, just like numerical identifiers (e.g. age) (Martnez et al., 2012).

Because the ”utility of textual information is closely related to the preservation of its meaning” (Martnez et al., 2012), semantic knowledge is often incorporated to find a reasonable balance between data quality and user privacy. For instance, Martnez et al. (2012) make use of the semantic knowledge encoded in the ontology WordNet. By merging semantically similar values into indistinguishable groups, the *k*-anonymity property is satisfied. This way, the authors' technique preserves the meaning of the textual values and thus a high data utility.

Another approach that relies on external knowledge bases (e.g. WordNet and SNOMED-CT, a healthcare knowledge base) is described by Snchez et al. (2014). The authors propose an automatic sanitization method applicable to textual data such as electronic healthcare records. The semantic knowledge is used to generalize private attributes (e.g. diseases) as well as semantically related concepts (e.g. symptoms) while still preserving a good data utility.

Li and Qin (2017) also focus on the anonymization of medical text records for subsequent sharing. Their approach begins with the clustering of the textual records into several groups according to the health-related information. Then, the sensitive aspects such as the age of the patient, date of the medical care, or the hospitalization are enumerated in (separate) lists so that the exact values cannot be associated with the individuals. This value-enumeration guarantees a proper degree of anonymization while at the same time leading to less information loss than the altogether masking of the aspects.

Bäumer et al. (2017) apply natural language processing techniques to automatically detect privacy breaches in reviews published on PRWs. They found out that the combination of private information disclosed in the natural language reviews and the provided meta data can enable de-anonymization. The fact that this is possible, even though websites like Jameda.de have already implemented mechanisms to detect severe privacy breaches such as personal names, makes it more important to figure out ways of how to protect users' privacy online.

Luo et al. (2016) present an approach to extract structured information from unstructured medical records. Their main goal is the automatic creation of a semi-structured database, not privacy protection itself. Nevertheless, the authors highlight the importance to protect data privacy and apply measures to this end. They divide the extracted medical data into separate modules so that identification across these modules is impossible.

There exist further studies that investigate privacy protection of unstructured data such as clinical text (Meystre et al., 2010, 2014; Gardner and Xiong, 2008). Even neural networks are used for this purpose: Dernoncourt et al. (2017) trained a bidirectional long short-term memory neural network on the de-identification of patient notes. Their network outperforms other state-of-the-art systems and succeeds in modeling the variations of natural language.

5 DISCUSSION

Most privacy-preserving approaches deal with larger data sets that comprise information about many in-

dividuals. When it comes to protecting the privacy of users on PRWs, it would be necessary to detect breaches *before* the users decide to publish their reviews. This is an additional challenge since an application has to detect breaches individually in every single review and not in a larger collection of reviews. Therefore, models that protect privacy relative to a batch of data points are not applicable in this usage.

There seems to be an inconsistency between a person's assessment of the value to protect one's privacy and the willingness to give it up and sell it (McDonald and Cranor, 2010; Cofone, 2017). This phenomenon is referred to as *privacy paradox* (Cofone, 2017). Users' motivations should be investigated so that implementations can be fit better to the needs and attitudes of the users. Min and Kim (2015) provide an overview of the costs and benefits that users take into account when considering (the extent of) using a social networking site. The perceived security and control they have of and over their data is a significant factor in the cost-benefit calculation. Furthermore, Dankar and El Emam (2012) provide insights into patients' concerns and uncertainties, especially when they are unsure about the further usage of their private information. Hence, further research should investigate the various privacy concerns of different users. Taking these into account for development, new applications are more likely to be accepted and adopted.

6 CONCLUSION

When patient data is published or shared, measures are taken to protect the individuals' privacy. This is important since from an ethical and social viewpoint, patients generally assume that their data is treated confidentially.

In the field of privacy protection, research should not only focus on structured data that is easier to analyze and anonymize. The increasing amount of natural language data that is available on and mined from the Web makes more research in this direction much-needed. Measures for privacy protection in general first have to arrive at sensible definitions of the concepts privacy and sensitive attributes. Concerning unstructured natural language data, the additional challenge of detecting the private attributes in the free text further complicates the problem. Especially user-generated content is still a type of data that is too seldom focalized by studies.

On the basis of the observations and research gaps presented above, we are planning to further investigate the breaches that users commit online on PRWs. Our goal is to build a system that automatically detects privacy breaches that users produce online. We are not aware of a working software solution that serves this purpose. The intended software would raise users' awareness concerning their privacy revelations by highlighting the infringements *on-line* and *before* the user decides to publish his or her post. This way, accidental disclosure of sensitive information could be prevented.

ACKNOWLEDGEMENTS

This work was partially supported by the German Research Foundation (DFG) within the Collaborative Research Centre On-The-Fly Computing (SFB 901).

REFERENCES

Ayala-Rivera, V., McDonagh, P., Cerqueus, T., and Murphy, L. (2014). A systematic comparison and evaluation of k-anonymization algorithms for practitioner. *Transactions on data privacy*, 7(3):337–370.

Bäumer, F. S., Grote, N., Kersting, J., and Geierhos, M. (2017). Privacy matters: Detecting nocuous patient data exposure in online physician reviews. In *Proceedings of the 23rd International Conference on Information and Software Technologies, Communications in Computer and Information Science*, volume 756, pages 77–89, Druskininkai, Lithuania. Springer.

Brody, S. and Elhadad, N. (2010). Detecting salient aspects in online reviews of health providers. *AMIA Annual Symposium Proceedings*, 2010:202–206.

Cofone, I. N. (2017). A healthy amount of privacy: Quantifying privacy concerns in medicine. *Cleveland State Law Review*, 65:1–26.

Dankar, F. K. and El Emam, K. (2012). The application of differential privacy to health data. In *Proceedings of the 2012 Joint EDBT/ICDT Workshops*, pages 158–166. ACM.

Dernoncourt, F., Lee, J. Y., Uzuner, O., and Szolovits, P. (2017). De-identification of patient notes with recurrent neural networks. *Journal of the American Medical Informatics Association*, 24(3):596–606.

Duncan, G. and Lambert, D. (1989). The risk of disclosure for microdata. *Journal of Business and Economic Statistics*, 7(2):207–217.

El Emam, K. (2006). Data anonymization practices in clinical research: A descriptive study. *University of Ottawa*, pages 1–15.

El Emam, K. (2011). Methods for the de-identification of electronic health records for genomic research. *Genome medicine*, 3(4):25.

El Emam, K., Dankar, F. K., Issa, R., Jonker, E., Amyot, D., Cogo, E., Corriveau, J.-P., Walker, M., Chowdhury, S., Vaillancourt, R., Roffey, T., and Bottomley, J. (2009).

A globally optimal k-anonymity method for the de-identification of health data. *Journal of the American Medical Informatics Association*, 16(5):670–682.

Ellimoottil, C., Hart, A., Greco, K., Quek, M. L., and Farooq, A. (2013). Online reviews of 500 urologists. *The Journal of Urology*, 189(6):2269–2273.

Emmert, M., Maryschok, M., Eisenreich, S., and Schffski, O. (2009). Arzt-Bewertungsportale im Internet Geeignet zur Identifikation guter Arztpraxen? *Das Gesundheitswesen*, 71(4):e18–e27.

Emmert, M., Meier, F., Heider, A.-K., Drr, C., and Sander, U. (2014). What do patients say about their physicians? an analysis of 3000 narrative comments posted on a german physician rating website. *Health Policy*, 118(1):66–73.

Emmert, M., Sander, U., Esslinger, A. S., Maryschok, M., and Schffski, O. (2012). Public reporting in germany: the content of physician rating websites. methods of information in medicine. *Methods of information in medicine*, 51(2):112–120.

Emmert, M., Sander, U., and Pisch, F. (2013). Eight questions about physician-rating websites: A systematic review. *Journal of Medical Internet Research*, 15(2):e24.

Fung, B. C. M., Wang, K., Chen, R., and Yu, P. S. (2010). Privacy-preserving data publishing: A survey on recent developments. *ACM Computing Surveys*, pages 1–55.

Gal, T. S., Chen, Z., and Gangopadhyay, A. (2008). A privacy protection model for patient data with multiple sensitive attributes. *International Journal of Information Security and Privacy*, 2(3):28–44.

Ganu, G., Kakodkar, Y., and Marian, A. (2012). Improving the quality of predictions using textual information in online user reviews. *Information Systems*, 38(1):1–15.

Gardner, J. and Xiong, L. (2008). Hide: an integrated system for health information de-identification. In *Proceedings of the 21st IEEE International Symposium on Computer-Based Medical Systems*, pages 254–259.

Han, M., Li, L., Xie, Y., Wang, J., Duan, Z., Li, J., and Yan, M. (2018). Cognitive approach for location privacy protection. *IEEE Access*, 6:13466–13477.

HIPAA (1996). The health insurance portability and accountability act of 1996 (HIPAA). *Department of Health and Human Services, United States of America*.

Kadry, B., Chu, L. F., Kadry, B., Gammas, D., and Macario, A. (2011). Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. *Journal of Medical Internet Research*, 13(4):e95.

Kersting, J., Bäumer, F. S., and Geierhos, M. (2019). In reviews we trust: But should we? experiences with physician review websites. In *Proceedings of the 4th International Conference on Internet of Things, Big Data and Security*, pages 147–155. SciTePress - Science and Technology Publications.

Kim, S., Sung, M. K., and Chung, Y. D. (2014). A framework to preserve the privacy of electronic health data streams. *Journal of Biomedical Informatics*, 50:95–106.

Li, F., Chen, J. Y., Zou, X., and Liu, P. (2010). New privacy threats in healthcare informatics: When medical records join the web. In *9th International Workshop on Data Mining in Bioinformatics*, BIOKDD 2010, pages 1–4.

Li, X.-B. and Qin, J. (2017). Anonymizing and sharing medical text records. *Information Systems Research*, 28(2):332–352.

Luo, L., Li, L., Hu, J., Wang, X., Hou, B., Zhang, T., and Zhao, L. P. (2016). A hybrid solution for extracting structured medical information from unstructured data in medical records via a double-reading/entry system. *BMC Medical Informatics and Decision Making*, 16(1):114.

Machanavajjhala, A., Gehrke, J., Kifer, D., and Venkitasubramaniam, M. (2006). l-diversity: Privacy beyond k-anonymity. In *22nd International Conference on Data Engineering (ICDE'06)*, pages 1–12. IEEE.

Martnez, S., Snchez, D., and Valls, A. (2011). Evaluation of the disclosure risk of masking methods dealing with textual attributes. *International Journal of Innovative Computing, Information and Control*, 8(7(A)):4869–4882.

Martnez, S., Snchez, D., Valls, A., and Batet, M. (2012). Privacy protection of textual attributes through a semantic-based masking method. *Information Fusion*, 13(4):304–314.

McDonald, A. M. and Cranor, L. F. (2010). Americans' attitudes about internet behavioral advertising practices. In *Proceedings of the 9th Annual ACM Workshop on Privacy in the Electronic Society*, WPES '10, pages 63–72. ACM.

McLennan, S., Strech, D., Meyer, A., and Kahrass, H. (2017). Public awareness and use of german physician ratings websites: Cross-sectional survey of four north german cities. *Journal of Medical Internet Research*, 19(11):e387.

Mendes, R. and Vilela, J. P. (2017). Privacy-preserving data mining: methods, metrics, and applications. *IEEE Access*, 5:10562–10582.

Meystre, S. M., Friedlin, F. J., South, B. R., Shen, S., and Samore, M. H. (2010). Automatic de-identification of textual documents in the electronic health record: a review of recent research. *BMC medical research methodology*, 10(1):70.

Meystre, S. M., scar Ferrndez, Friedlin, F. J., South, B. R., Shen, S., and Samore, M. H. (2014). Text de-identification for privacy protection: A study of its impact on clinical text information content. *Journal of biomedical informatics*, 50:142–150.

Min, J. and Kim, B. (2015). How are people enticed to disclose personal information despite privacy concerns in social network sites? the calculus between benefit and cost. *Journal of the Association for Information Science and Technology*, 66(4):839–857.

Okike, K., Peter-Bibb, T. K., Xie, K. C., and Okike, O. N. (2016). Association between physician online rating

and quality of care. *Journal of Medical Internet Research*, 18(12):e324.

Pestian, J. P., Brew, C., Matykiewicz, P., Hovermale, D., Johnson, N., Cohen, K. B., and Duch, W. (2007). A shared task involving multi-label classification of clinical free text. In *Proceedings of the Workshop on BioNLP 2007: Biological, Translational, and Clinical Language Processing*, pages 97–104. Association for Computational Linguistics.

Samarati, P. and Sweeney, L. (1998). Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. *Technical Report SRI-CSL-98-04*, pages 1–19.

Shah, A. and Gulati, R. (2016). Privacy preserving data mining: Techniques classification and implications a survey. *International Journal of Computer Applications*, 137(12):40–46.

Snchez, D., Batet, M., and Viejo, A. (2013). Minimizing the disclosure risk of semantic correlations in document sanitization. *Information Sciences*, 249:110–123.

Snchez, D., Batet, M., and Viejo, A. (2014). Utility-preserving privacy protection of textual healthcare documents. *Journal of biomedical informatics*, 52:189–198.

Wagner, I. and Eckhoff, D. (2018). Technical privacy metrics: a systematic survey. *ACM Computing Surveys*, 51(3):57.

Yang, Y., Lutes, J., Li, F., Luo, B., and Liu, P. (2012). Stalking online: on user privacy in social networks. In *Proceedings of the second ACM Conference on Data and Application Security and Privacy*, pages 37–48. ACM.