

Social Network Analysis of Health Development in Indonesia

Agung Tika Wicaksono¹, Siti Mariyah²

STIS Polytechnic of Statistics

Jakarta, Indonesia

¹15.8471@stis.ac.id, ²sitimariyah@stis.ac.id

Abstract—The regular availability of up to date data for evaluating the government's policy is still lack. One of them is data for evaluating development in the health sector. Data are crucial for decision making and policy evaluation. Public perspective or sentiment to each program initiated by the government is a form of evaluation from the citizen. Therefore, we need a technique that can regularly supply data used for decision-maker to measure the success of development programs. Utilization of news can help overcome the cost and time in carrying out updates on development initiated and performed by the government. In this study, we developed an application which can collect health development-related online news, process and analyze them to reflect the public's perspective to the health development programs. Successfully collected 1,204 news articles where 117 articles from *detik.com*, 661 articles from *kompas.com* and 426 articles from *tempo.co*. A total of 500 from 1,204 news articles are used as training data for making Named Entity Recognition models. Sentiment analysis of research results shows that the sentiments of the public to the issue of equitable quality of health services, financial protection, and distribution of drugs and medical personnel are positive. The results of the study indicate that news can be used as an analytical material for evaluating development, the results of which are quite relevant to the results achieved.

Keywords— *named entity recognition, social network analysis, news, sentiment analysis, health development*

I. INTRODUCTION

Development is a planned and systematic effort by each and all elements of the nation to improve the situation by utilizing the resources owned effectively, efficiently, optimally and accountably, with the aim of improving the quality of life of people and society in a sustainable manner [1]. Health development is a Healthy Indonesia Program aimed at improving the health status and nutritional status of the community through community empowerment by providing financial protection and equitable health services [2]. The targets for health development include increasing maternal and child health and nutrition status, improving disease control, improving access and quality of basic and referral health services especially in remote, disadvantaged and border areas, increasing coverage of universal health services through the Healthy Indonesia Card (Kartu Indonesia Sehat) and management of the National Health Insurance System (BPJS),

meeting the needs of health workers, drugs and vaccines, and increasing the health system's responsiveness.

It is important to control whether the development carried out is on the target or not. To find this out, an assessment of the course of the development program is needed. The assessment that can be done is by monitoring and evaluating. In monitoring and evaluating, it is very likely that the data held to be used as material for monitoring and evaluation are not yet available from the relevant agencies in a short time. While the need for data to be used as an evaluation material for the development carried out becomes mandatory to be fulfilled. If the available data is faster, then an evaluation of development will be easier to do and vice versa. If all the data needed is still difficult to collect immediately, it is necessary to check the benefits and or difficulties of the development program through a sample survey to overcome the problem of data availability.

Evaluations conducted conventionally often lead to several problems, such as a considerable lag time between the time of the evaluation and the time of publication of the evaluation results. The problem that follows the long time lag is that the evaluation results are no longer relevant to the conditions at the time the publication of the evaluation results. Evaluations conducted in a conventional manner have also not been able to see the influence of figures, positions, organizations, and locations on health development in Indonesia. Therefore, we try to utilize online news to overcome these problems. It is caused online news contains the aspirations of the people represented by several figures both from government and from outside the government, in this case, aspiration related to health development. The availability of available online news continuously allows the evaluation results to be published as soon as possible to reduce the time lag as in conventional evaluations in general. The Named Entity Recognition (NER) model was also built to detect the influence of figures, positions, organizations, and locations on health development undertaken to strengthen the results of evaluations conducted using online news.

II. RELATED WORKS

There is a study to analyze public sentiment towards public health programs using Twitter. The research is based on evaluating and improving to facilitate the government gets feedback on the existing programs and policies. The method

used to produce sentiments is lexicon based. Tweets that have been collected are then performed data cleaning by removing some of the symbols contained in the tweet such as the @ sign, RT (Re-Tweet), # (hashtag), and URL. Tweets that are already clean then do preprocessing data in the form of changing words into lowercase, eliminating punctuation, and stopword removal. Irawan [4] used the rubric "Report Uncle Sripo" (*Lapor Mang Sripo*) in the Sriwijaya Post newspaper to assess public services by the city of Palembang. There are 16 short message service (SMS) used as analysis material. The method used is qualitative content analysis with a descriptive approach. The SMS is then categorized into seven categories for analysis purposes. Unfortunately, the process of classifying SMS is still done manually so that it will be difficult to do an analysis when dealing with large amounts of data. Hidayat [5] used 39 news articles from the Daily Analysis Newspaper to assess the 100 days of performance of the Mayor and Deputy Mayor of Medan for the period 2010-2015. The research used descriptive methods and applied content analysis to explore the information contained in each news article. News about the economy can also be used to analyze the economy in related countries [6] and to predict the impact of news sentiment on the stock market [7].

III. PROPOSED METHOD

This research examined whether health development programs in Indonesia from 2015 to 2019 have been going well or not. Before generating sentiment values for each news and opinion in the news, the pre-processing stage is carried out first. In the activities of the pre-processing stage such as separating words, deleting symbols, changing words into lowercase letters and deleting stopwords. The clean data are then checked whether the words in the clean data are contained in the dictionary of positive and negative words or not [8][9]. Then, we calculated the number of positive and negative words to get the value of sentiment.

In order to extract information about persons, positions, opinions, organizations, and locations in the news, we constructed the Named Entity Recognition (NER) using SpaCy library. We need a model for each of the related variables. The model was obtained by training 500 news articles. Training data were generated from the manual labeling stages using the Gate Developer application. The training data that generated at the labeling stage is manually stored in xml form. Then we have to convert the training data into json form so it can be used on SpaCy. Training using SpaCy was performed to obtain the NER model based on five entities, namely person, position, opinion, organization, and location. The accuracy of the model in extracting information is determined by the accuracy and accuracy of researchers in labeling. The amount of data used as training, as well as the number of iterations when conducting model formation training also determines the accuracy of the model.



Fig. 1. Research framework

A. Data Collection

In this study, we used two data collection techniques, namely literature study and crawling.

1) *Literature Study*: The aim was to dig deeper information regarding the indicators that will be analyzed in relation to health development in Indonesia. Information collected can be in the form of statistics, problems, phenomena that develop regarding related indicators, concepts and definitions used when labeling in making constructing NER models. Literature study in this study comes from the results of publications, relevant ministry websites, online media, and related research.

2) *Crawling*: News articles used as a training set, testing set, and analysis needs are crawled using robots made using Kapow. The process begins with installing Kapow on the researcher's computer. Then start Kapow by opening the Start Management Console until the RoboServer window opens. Next is filling the Kapow license. The license used by researchers is an official license originating from the Central Statistics Agency (BPS). After that, we designed the Kapow robot in Design Studio 9.6.2 [12]. The next step is to configure the database so that the crawling data can be stored directly in

the database. One robot can only be used for one news website. This is because the design of each news website is different between one news website and another news website. Not all news on online news websites is collected. The only news that corresponds to the health development sub-indicator keyword and published from 2015 to 2019 will be collected. The keywords that are used during the crawling process are *angka kematian ibu* (maternal mortality rate), *angka kematian bayi* (infant mortality rate), *kurang gizi* (malnutrition), *stunting*, *tuberkulosis* (tuberculosis), HIV, malaria, *darah tinggi* (high blood pressure), *obesitas* (obesity), *perokok* (smoker), *puskesmas* (community health centers), *rumah sakit* (hospital), *RSUD* (regional public hospital), *imunisasi dasar* (basic immunization), *SJSN* (National Social Security System), *obat* (drugs), and *vaksin* (vaccine). Successfully collected 1,204 news stories, of which 117 of them came from *detik.com*, 661 news came from *kompas.com* and 426 other news came from *tempo.co*. News is said to be clean because the news is free from duplicate news content. Duplication occurs during the process of crawling news. Duplicated news articles were marked with the exact same news content but have a different URL. The way to overcome this news duplication was to query the database using a distinct function to search for news with identical content that differs from one story to another. A total of 500 news from 1,204 news articles were used as training data for making the NER [10] model, where the labeling is done manually at Gate Developer [11].

B. Data Pre-processing

There are two different stages of pre-processing, namely pre-processing on the news that will be used for sentiment analysis and pre-processing on the news that will be used as a training set for the formation of the NER model. At the pre-processing stage for news that will be analyzed sentiment, we conducted some tasks such as deletion of symbols and punctuation; no standardization of words was done because in general words contained in the news are standard words; transforming words to lowercase; the division of text in the form of sentences or paragraphs into certain parts of the section (tokenization); deletion of words that are not needed in the analysis (stopword).

Meanwhile, in the news preprocessing stage which will be used as training data for constructing the NER model, the pre-processing stages are carried out by deletion of words and symbols that exist right before and after this news, for example news sources, news source locations, news writers and so on; removing advertisements that are embedded in the news articles; correcting typos and punctuation errors in the news, for example the size of letters and punctuation.

C. Data Processing

In the processing stage, two different processing stages were performed, namely processing to produce the sentiment value and to obtain the results of entity extraction by NER model. Processing to produce sentiment value began with the process of news article tokenization. Then, sentiments were weighted from tokenized news articles. Weighting is done

using a positive word dictionary and a negative word dictionary that is matched to each word in the news so that the sentiment value for each word is obtained [8][9]. The positive and negative word dictionary had been adjusted based on these research needs, for example, there are some words that should not need to be included in positive words and negative word dictionaries but were included. After getting sentiment scores for each word, we summed those for overall news sentiment values. This method was known as the lexicon method. If a word matches the words in the positive word dictionary, then the positive value counter for the news is increased by one. Whereas if a word matches the word in the negative word dictionary, the negative value counter for the news is increased by one. A news article was categorized as having positive sentiment if the counter of the positive value is greater than the counter of the negative value, vice versa. Whereas a news article was categorized as having neutral sentiment if the counter of positive value and counter of a negative value is the same.

We constructed a NER model to extract target entities in every news articles. We applied supervised learning in developing NER model, by using 500 articles for the training set and 50 articles for the testing set. We defined five entities which should be extracted from each news article. We labeled 500 training manually using GATE. We labeled “person” for each word stating person, labeled “position” for each word stating position of person in the organization, labeled “organization” for each word stating cooperation groups between people held to achieve certain goals, both government and non-government contained in the news, labeled “location” for each word stating countries, provinces, districts, cities, and sub-districts listed in the news, and labeled “quote” for each word stating words or thoughts of people, can be either direct or indirect sentences contained in the news. The produced NER models were used to recognize and extract target entities for all news articles.

D. Analysis Technique

The analytical methods applied were analytical sentiment analysis, content-based analysis, and social network analysis. Sentiment analysis was carried out to assess the sentiment of news whether it is positive, negative, or neutral. The content-based analysis was applied to look deeply into the information contained in the news, including social phenomena that are often reported. Social network analysis was performed to capture the relationship between the entities obtained from the NER model that has been built.

IV. RESULT AND DISCUSSION

We evaluated health development through indicators established in the 2015-2019 National Medium-Term Development Plan (*Rencana Pembangunan Jangka Menengah Nasional/ RPJMN*). There are four indicators of health development, namely health and nutrition status, control of communicable and non-communicable diseases, equity and quality of health services, and financial protection, availability, distribution and quality, and health resources. Due to page limitations, we only discuss one sub-indicator for each indicator.

TABLE I. KEYWORD AND SENTIMENT RESULTS FOR EVERY HEALTH AND NUTRITIONAL STATUS SUBINDICATORS

No	Health and Nutrition Status		
	Subindicator	Keyword	Sentiment
1	Maternal mortality rate per 100,000 live births	maternal mortality rate	Negative
2	Infant mortality rate per 1,000 live births	infant mortality rate	Negative
3	The prevalence of malnutrition (underweight) in children under five (percent)	malnutrition	Negative
4	The prevalence of malnutrition (underweight) in children under five (percent)	stunting	Positive

A. Indicators of Health and Nutrition Status

Based on table I, it can be seen that three of the four subindicators of health and nutrition status have negative sentiments.

Based on Fig. 2, as many as 52.08% of the total 98 news that discussed the Maternal Mortality Rate had negative sentiments. While 44.90% have positive sentiment and another 3.06% have a neutral sentiment. The dominance of negative sentiment towards positive sentiment is in line with the facts on the ground which show that the maternal mortality rate in Indonesia is still relatively high. In Indonesia, the Maternal Mortality Rate surpasses 305 per 100,000 live births. The maternal mortality rate is still quite high when compared to other countries in ASEAN (Association of Southeast Asian Nations), which ranges from 40-60 per 100,000 live births. A lot of news discusses the factors that cause maternal death, such as early marriage, knowledge of pregnant women, bleeding, heart disease, and so forth. This also contributed to negative news sentiment. Here are some examples of extracted quotes that have negative sentiments:

- “Gubernur Jawa Tengah mengaku galau dengan tingginya Angka Kematian Ibu (AKI) dan Angka Kematian Bayi (AKB) di Jawa Tengah yang terbilang tinggi”
- “mengatakan, rendahnya pengetahuan ibu hamil menjadi salah satu faktor tingginya angka kematian ibu (AKI) dan angka kematian bayi (AKB)”
- “Perkawinan usia dini memicu tingginya angka kematian ibu”.

Pie Chart of Sentiment Results for Maternal Mortality Rate

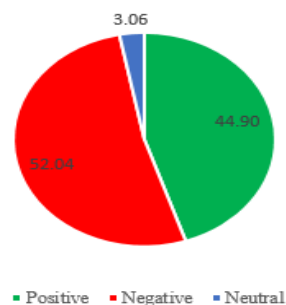


Fig. 2. Pie chart of sentiments for maternal mortality rate

Bar Chart of The Extraction Results of Entity Location for HIV

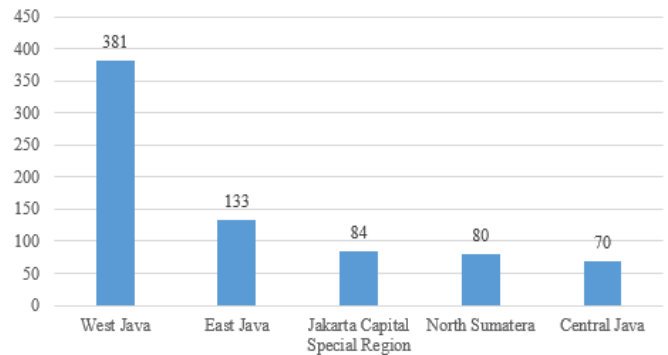


Fig. 3. Bar chart for the extraction of location entities of HIV

TABLE II. KEYWORD AND SENTIMENT RESULTS OF SUBINDICATORS IN CONTROLLING INFECTIOUS AND NON-COMMUNICABLE DISEASE

No	Controlling Infectious And Non-Communicable Disease		
	Subindikator	Keyword	Sentiment
1	Prevalence of tuberculosis (TB) per 100.00 population	tuberculosis	Negative
2	HIV prevalence (percent)	HIV	Negative
3	Number of districts / cities achieving malaria elimination	malaria	Negative
4	Prevalence of high blood pressure (percent)	high blood pressure	Negative
5	The prevalence of obesity in people aged 18 years and over (percent)	obesity	Negative
6	Smoking prevalence is less than 18 years old	smoker	Negative

B. Indicators of Control of Contagious and Non-Communicable Diseases

Based on table II, it shows that all sub-indicators controlling infectious and non-communicable diseases possess negative sentiments.

Based on the results of location entity extraction using the NER model that has been made in Fig. 3, the results of the top five locations that have links with other locations on the news with the keyword HIV namely West Java (381 connections), East Java (133 connections), Jakarta Capital Special Region (84 connections), North Sumatera (80 connections), and Central Java (70 relationships). Meanwhile, based on the extraction results of entity person and entity position, the entities with the most frequent occurrences include:

- Rini (146 times), penderita HIV.
- Jusuf Kalla (115 times), Wakil Presiden Republik Indonesia.
- Daniel (72 times), Brand Manager Kondom Sutra.

The persons above are persons that are often published in the news and play a role in providing information related to HIV.

TABLE III. KEYWORD AND SENTIMEN RESULTS OF SUBINDIKATOR EQUITY AND QUALITY OF HEALTH SERVICES

No	Equity and Quality of Health Services		
	Subindicator	Keyword	Sentiment
1	The number of sub-districts that have at least one community Health centers that is certified by national accreditation	community health centers	Positive
2	Number of districts / cities that have one RSUD that is certified by national accreditation	hospital, regional public hospital	Negative
3	Percentage of districts / cities that reach 80 percent complete basic immunization in infants	basic immunization	Neutral

C. Indicators of Equity and Quality of Health Services

Based on table III, it describes three of the four sub-indicators of health and nutrition statuses have negative sentiments.

Referring to Fig 4, it can be seen that as many as 59.42% of the total 313 news related to Puskesmas have positive sentiments, 36.10% have negative sentiments and the remaining 4.47% have neutral sentiments. The results of this sentiment are fairly in accordance with the conditions on the ground, where health care facilities in puskesmas are already quite capable, especially in border areas. Even the Regional Director of the World Health Organization of the South-East Asia Region (WHO SEARO), Poonam Khetapral Singh, expressed his appreciation for efforts to improve the quality of health centers in Indonesia. This was conveyed at the WHO Executive Board Program, Budget and Administration Committee (PBAC) meeting in Geneva, Switzerland, which was held on January 21-23, 2019. In terms of health development, it was also quite massive. A total of 256 puskesmas were built at the border throughout 2018. Most of the news discussed the improvement of puskesmas services and the construction of puskesmas. This is one of the causes of news about the health center to have a positive sentiment.

Here are some examples of the extraction results of quotes that have positive sentiment:

- “mengatakan akan membangun enam puskesmas baru yang memiliki fasilitas kamar inap yang cukup”.

Pie Chart of Sentiment Results for Community Health Centers

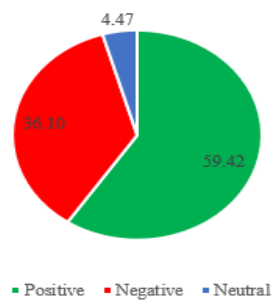


Fig. 4. Pie chart of sentiments for community health centers

TABLE IV. KEYWORD AND SENTIMEN RESULTS OF SUBINDICATORS OF FINANCIAL PROTECTION, AVAILABILITY, DISTRIBUTION AND QUALITY AND HEALTH RESOURCES

No	Financial Protection, Availability, Distribution and Quality and Health Resources		
	Subindicator	Keyword	Sentiment
1	Percentage of healthcare SJSN (percent)	National Social Security System	Positive
2	The minimum number of Puskesmas has five types of health workers	community health enters	Positive
3	Percentage of class C district / city hospitals that have seven specialist doctors	hospital, regional public hospital	Negative
4	Percentage of drug and vaccine availability at the community health enters	drugs, vaccine	Positive
5	Percentage of drugs that meet the requirements	drugs	Positive

- “menjamin fasilitas pada Puskesmas layak, tak kalah dengan fasilitas kesehatan milik swasta”.
- “Puskesmas kita sudah mewah, dokternya juga sudah lengkap”

D. Indicators of Financial Protection, Availability, Distribution and Quality and Health Resources

Table IV explains that four of the five sub-indicators are financial protection, availability, distribution and quality of drugs and health resources.

Based on the extraction results of entity person and entity position, the entity with the most frequent occurrences is shown, among others:

- Jokowi (387 times), Presiden Republik Indonesia.
- Debora (277 times), bayi korban meninggal di RS Mitra Keluarga Kalidere.
- Sukiyat (224 times), Presiden Komisaris PT Kiat Mahesa Wintor Indonesia.

The persons above are persons that are often published in the news and play a role in providing information related to the hospital.

Colorful dots in Fig. 5 represents the organization while a straight line connecting the dots indicates that the dots have a relationship. The connection here is because the organization appeared on the same news. Referring to Figure 5, the results of the top five organizations that have relationships with other organizations related to hospitals are BPJS Health (1058 relationships), BPJS (985 relationships), government (629 relationships), Mitra Hospital (345 relationships), and the Ministry of Health (296 relationships).

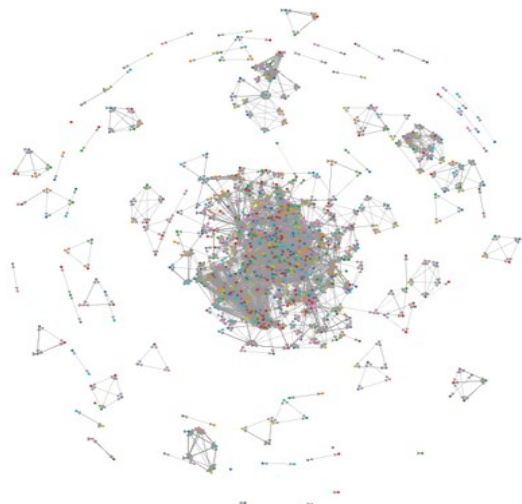


Fig. 5. Social network analysis of organization entities of hospital

V. EVALUATION

An evaluation of sentiment results was conducted by comparing the sentiments measured in this study with the results of the 2015-2019 RPJMN evaluation published by the National Development Planning Agency. The results of the publication are said to be positive if the aggregate of the indicators is more positive and vice versa, the results of the publication are said to be negative if the aggregate of the indicators is more negative. While the results of the publication are neutral if the aggregate indicators are positive and negative the same amount. The following is a comparison table between the results of publications with the results of research sentiments:

Table V shows that two of the four indicators of research findings are in agreement with the results of the publication, namely in assessing the last two indicators of health development. While the two initial indicators have different values between the results of the study with the results of the publication. This difference occurs because most of the news that discusses the two initial indicators are more inclined to discuss the impact caused and the causes of the problems in

TABLE V. COMPARISON OF ASSESSMENT OF PUBLICATION RESULTS AND RESEARCH

No.	RPJMN Indicators		
	Indikator	Publication	Research
1	Health and nutrition status	Negative	Neutral
2	Controlling Infectious And Non-Communicable Disease	Negative	Neutral
3	Equity and Quality of Health Services	Positive	Positive
4	Financial Protection, Availability, Distribution and Quality and Health Resources	Positive	Positive

these indicators, not to the data that actually shows a fairly good trend.

VI. CONCLUSION

We have succeeded in building methods from the data collection stage to evaluation. This included building a NER model to identify important entities that can be utilized to evaluate health development. This research proved that evaluating health development can be done by utilizing online news. Online news can be used as material for analysis to assess development, because the results are quite relevant to the original situation. The results of sentiment analysis show that the development in the health sector in Indonesia is quite good. Nine out of ten sub-indicators related to health and nutritional status as well as controlling infectious and non-communicable diseases received negative sentiment.

REFERENCES

- [1] Nasional, K. P. P. (2014). Rencana Pembangunan Jangka Menengah Nasional 2015-2019. Jakarta: Kementerian Perencanaan Pembangunan Nasional/Badan Perencanaan Pembangunan Nasional.
- [2] Kementerian Kesehatan, R. I. (2015). Rencana strategis kementerian kesehatan tahun 2015-2019. Jakarta: Kementerian Kesehatan RI.
- [3] Ariwibowo, A. S. (2018). Analisis Sentimen Publik pada Program Kesehatan Masyarakat menggunakan Twitter Opinion Mining. Seminar Nasional Informatika Medis. Yogyakarta: Universitas Pembangunan Nasional "Veteran".
- [4] Irawan, P. (2017). Analisis Opini Publik Tentang Kualitas Pelayanan Publik Pemerintah Kota Palembang dalam Rubrik "Lapor Mang Sripo" pada Surat Kabar Sriwijaya Post (Doctoral dissertation, UIN RADEN FATAH PALEMBANG).
- [5] Hidayat, T. W. (2016). Analisis Berita Kesehatan di Media Massa terhadap Pelayanan Publik. JURNAL SIMBOLIKA: Research and Learning in Communication Study, 1(2).
- [6] Nurwidiantoro, A. (2016, October). Sentiment analysis of economic news in Bahasa Indonesia using majority vote classifier. In 2016 International Conference on Data and Software Engineering (ICoDSE) (pp. 1-6). IEEE.
- [7] Shah, D., Isah, H., & Zulkernine, F. (2018, December). Predicting the Effects of News Sentiments on the Stock Market. In 2018 IEEE International Conference on Big Data (Big Data) (pp. 4705-4708). IEEE.
- [8] Liu, B., Hu, M., & Cheng, J. (2005, May). Opinion observer: analyzing and comparing opinions on the web. In Proceedings of the 14th international conference on World Wide Web (pp. 342-351). ACM.
- [9] Wahid, D. H., & Azhari, S. N. (2016). Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity. IJCCS (Indonesian Journal of Computing and Cybernetics Systems), 10(2), 207-218.
- [10] Choi, J. D., Tetreault, J., & Stent, A. (2015). It depends: Dependency parser comparison using a web-based evaluation tool. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers) (Vol. 1, pp. 387-396).
- [11] H. Cunningham, V. Tablan, A. Roberts, K. Bontcheva (2013) Getting More Out of Biomedical Documents with GATE's Full Lifecycle Open Source Text Analytics. PLoS Comput Biol 9(2): e1002854. doi:10.1371/journal.pcbi.1002854
- [12] Kofax. Kofax RPA. Access April 2019. <https://www.kofax.com>