



Extraction of measurements from medical reports

El Youssefi Ahmed, Abdelahad Chraibi, Julien Taillard, Ahlame Begdouri

► To cite this version:

El Youssefi Ahmed, Abdelahad Chraibi, Julien Taillard, Ahlame Begdouri. Extraction of measurements from medical reports. 10ème conférence Francophone en Gestion et Ingénierie des Systèmes Hospitaliers, GISEH2020, Oct 2020, Valenciennes, France. hal-03229520

HAL Id: hal-03229520

<https://hal.science/hal-03229520>

Submitted on 19 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Extraction of measurements from medical reports

EL YOUSSEFI Ahmed ¹, CHRAIBI Abdelahad ², *TAILLARD Julien* ², BEGDOURI Ahlame ³

¹ LSIA FST Fès, B.P. 2202 – Route d'Imouzzer - FES – MAROC, +212 (0) 653 17 31 94, ahmed.elyoussefi@usmba.ac.ma

² Alicante, 50 rue Philippe de Girard - 59113 SECLIN, (+33) 3 28 55 92 50, abdelahad.chraibi@alicante.fr, julien.taillard@alicante.fr

³ LSIA FST Fès, B.P. 2202 – Route d'Imouzzer - FES – MAROC, +212 (0) 535 60 80 14, ahlame.begdouri@usmba.ac.ma

Abstract. A patients' medical record represents their medical history and enclose interesting information about their health status within written reports. These reports usually contain measurements (among other information) that need to be reviewed before any new medical intervention, since they might influence the medical decision regarding the types of drugs that are prescribed or their dosage. In this paper, we introduce a method that extracts measurements automatically from textual medical discharge summaries, admission notes, progress notes, and primary care notes. We don't distinguish between reports belonging to different services. For doing so, we propose a system that uses Grobid-quantities to extract value/unit pairs, uses generated rules from analysis of medical reports and text mining tools to identify candidate measurements. These candidates are then classified using a Long Short Term Memory (LSTM) network trained model to determine which is the corresponding measurement to the value/unit pair. The results are promising: 95.13% accuracy, a precision of 92.38%, a recall of 94.01% and an F1 score of 89.49%.

Keywords: Measurement, Medical report, Natural Language Processing, Long Short Term Memory (LSTM), Conditional Random Fields (CRF).

Introduction

A patient's medical record, "retraces the episodes of a person's illness and care path" [Comité éditorial pédagogique de l'UVMaF, 2012], it is considered as an "intermediary between the patient and the medical world, it contains confidential information and a lifelong medical journey" of a patient [Hecquet, 2012]. These information provide a link of communication between health professionals (doctors, laboratory managers, pharmacists, etc.), and summarize the results of medical tests performed, physician observations, recommendations for the patient or for another stakeholder in the health sector.

Medical reports are usually written in free texts, that differ from a health actor to another, they are full of useful information about a patient. Most studies agreed that human review of text to is too time-consuming and labor intensive to be achievable, which require the implementation of methods to extract information automatically [Ford et al., 2016]. One of these information are measurements that might be:

- Medical test results
- Medicament dosage and frequency
- Information related to a patient (age, weight ...)
- ...

The extraction of such information (including measurements) may be useful for:

- Identifying the patient's response to prescribed drugs (change of medication or doses, between one report and another for the same patient in a given period).
- Identifying collateral effects of drugs or possible inter drugs effects (increasing/decreasing effectiveness).
- Identifying patients at risk of certain pathologies based on medical tests results. Early detection will allow physicians to take preventive actions to reduce the impacts of these pathologies or to cure them totally.
- Ensure interoperability between systems that use Electronic Health Records (EHRs) by sharing comprehensive view of a patient's information that takes the information from both systems into account. [Resnik et al., 2008]

In this work we propose a system that allows the extraction of measures contained in a given medical report, written in free text.

1. Previous work

A method have been proposed by [Mandhan, 2015] for the extraction of numerical attributes and values from patient discharge reports, and the association between them. The proposed solution tries to solve the problem in two steps. The first step is to detect the extraction of numerical attributes and values, which are developed as a named entity recognition (NER) model using the Stanford CoreNLP library. The second step is to correctly associate attributes with values, which are developed as a relationship extraction module in the Apache cTAKES Framework. The Stanford CoreNLP model for NER was integrated as a cTAKES pipeline component and used in the relationship extraction module. The conditional random field algorithm (CRF) was used for NER and support vector machine(SVM) for relationship extraction.

[Elise et al., 2015] addressed the problem of detecting and extracting numerical values from patients' narrative documents and associating them with corresponding "themes" and units. They opted for the use of a categorization supervised by conditional random fields for the detection of segments (themes, quantities and units). A system based on rules was used for the association of these segments in order to build semantically significant sequences.

The steps proposed in this work are:

- Data preprocessing: the segmentation into words of the corpus and the attribution of morphosyntactic labels.
- An additional annotation of documents with semantic resources is performed to remove the numerical values corresponding to the phone numbers, clinical identifiers ...
- A CRF model is responsible for detecting and extracting numerical values possibly related to the themes covered as well as the units.
- The association of extracted elements (theme T, numerical value N, unit U) in order to construct coherent and semantically significant sequences. This step is done with a rule-based approach, in which a set of models allow these elements to be associated with each other is used.
- The final step in this method is the evaluation of the baseline data provided by the physician.

A dynamic and independent approach to the field, to extract numerical values described from clinical documents written in the German language was proposed in [Zubke, 2017]. It does not use manually defined rules or ontologies or nomenclatures unlike other systems. This system considers the problem of extracting information as a problem of a text classification. Therefore, it uses machine learning to identify crucial context characteristics of a subject-specific numerical value through a given set of sample sentences. This reduces manual effort in selecting appropriate sample sentences. The context characteristics of a numerical value are described by term frequency vectors that are generated by several document segmentation procedures. Because of this simultaneous segmentation approach, there may be more than one context vector for a numerical value. In this case, the context vector with the greatest classification confidence is chosen.

The cited work have inspired us with the process of developing our system, especially the rule-based approach used in [B. Elise et al., 2015], even if in our context we are not trying to attribute a theme to a given quantity/unit pair, but instead we are trying to find the corresponding measurement. What is new within our approach is the use of semantic and syntactic analysis to identify token interrelations, and the use of neural network based classification.

2. Material and method

The used data consisted of 54 medical reports regarding patients, written in French language, and in free text format. All the reports have been anonymized.

The following steps were then conducted.

2.1 Preprocessing

The input document/text is preprocessed using regular expressions. For example, we separate text and quantities (numbers), we rewrite all decimals to use point as decimal separator, we rewrite some quantities (for example : 5kg340 will be 5.340 kg).

2.2 Grobid quantities training

The reports were then annotated using TEI format and used to train grobid quantities models. Grobid Quantities is a module of GROBID that stands for GeneRation Of Bibliographic Data. It allows "recognition in textual documents of any expression of measurements" [P. Lopez, 2019]. The role of Grobid-Quantities in our solution is to identify quantity/unit pairs, or quantity alone in the case of unit-free measurements.

2.3 Rules definition

These reports were then splitted into sentences using CoreNLP and only the sentences containing at least one measure were considered. The result was 667 sentences, with the morphosyntactic tag annotation related to each token (NC, N, ADJ, P, ADV, ...). For the prepositions we concatenate it with the ":" followed by the actual preposition, for example : **P:de** or **P:à**. Since we noticed that this preposition was considerably used and we opted for this solution to differentiate between them.

Based on the morphosyntactic tags of quantity token and unit token, we have established a JSON file containing a set of rules to identify tags of tokens representing measurement candidates.

We used 3 types of annotations for the morphosyntactic tags:

":u": indicates that the corresponding token is a unit.

":q": indicates that the corresponding token is a quantity.

":me": indicates that the corresponding token is a measure.

Each rule is made up of two parts

Match: Specify the morphosyntactic tag of quantity and unit

Search: indicates possible sequences of the measurement.

We have two types of rules:

2.3.1 Rules for measures with quantity only:

These rules correspond to measures in which only the quantities are present

Here's an example of the entry of this JSON file:

```
{
  "match": "NC:q",
  "chercher": "NC:me|ADJ:me|NC:me#NC:me|P:en|NC:me|NC:me#NC:me|P:à#NC:me|ADJ:me"
}
```

Figure 1. example of json rule for a measurement without unit

In this example, this rule can be interpreted as follows: If the morphosyntactic tag of the corresponding token of the quantity is equivalent to **NC**, we will look for one of the following token sequences (separated by # in the json file):

```

NC:me ADJ:me NC:me
NC:me P:en NC:me NC:me
NC:me P:à
NC:me ADJ:me

```

Figure 2. interpretation of the sample rule

Any correct sequence will be considered as a measurement candidate. The presence of **"me"** indicates that this word will be considered as a measure.

2.3.2. Rules for measurements with quantities and units:

These rules correspond to measures that are expressed using a quantity/unit pair.

```

{
  "match": "N:q|ADJ:u",
  "chercher": "N:me#NPP:me|V|P:à#V:me|ADJ:me|P:à#ADJ:me#NC:me|ADJ:me#NPP:me|P:à"
}

```

Figure 3. example of json rule for a measurement with a unit

This example could be interpreted as follow : if the morphosyntactic tag of the corresponding token of the quantity is equivalent to **N** and that of the unit is **ADJ**, we will look for one of the sequences of words contained in the "chercher" (means search in French) part.

In this example if the chosen rule of the system is **NPP:me V P:à**. For example, the measure will be returned to the corresponding word **NPP** only, since it is the only one that is annotated by **"me"**.

Based on these rules, we have developed a tool that generates the candidate measurement for each quantity or quantity/unit pairs. The result was 2876 candidates for the whole sentences, that have been annotated to determine if the candidate is the correct measure or no.

2.4 LSTM models training

LSTMs are a special type of RNN, capable of learning long-term relations. They are designed to evoke information for long periods of time [Hochreiter, Schmidhuber, 1997].

A measure in our corpus is part of a sentence, which is a sequence of words. So as to identify which measure is related to a quantity/unit pair, we must take into account the pair itself, including all of the previous (or next) words, and other previous measures.

This sequential aspect that led us to opt for the use of LSTMs in our proposed method.

We splitted our corpus into 2 sets:

We used 2158 annotated sentences for training, which is $\approx 75\%$ of the whole annotated sentences.

And 718 annotated sentences for test, which is $\approx 25\%$ of the whole annotated sentences.

We used the implementation of LSTM proposed by Deep Learning for JAVA (DL4J). This library will be also used for prediction using trained models within the system.

The used version of DL4J is 0.9.1

3. System structure

Our proposed system, as illustrated in Figure 4, consists of seven phases. They are successively described.

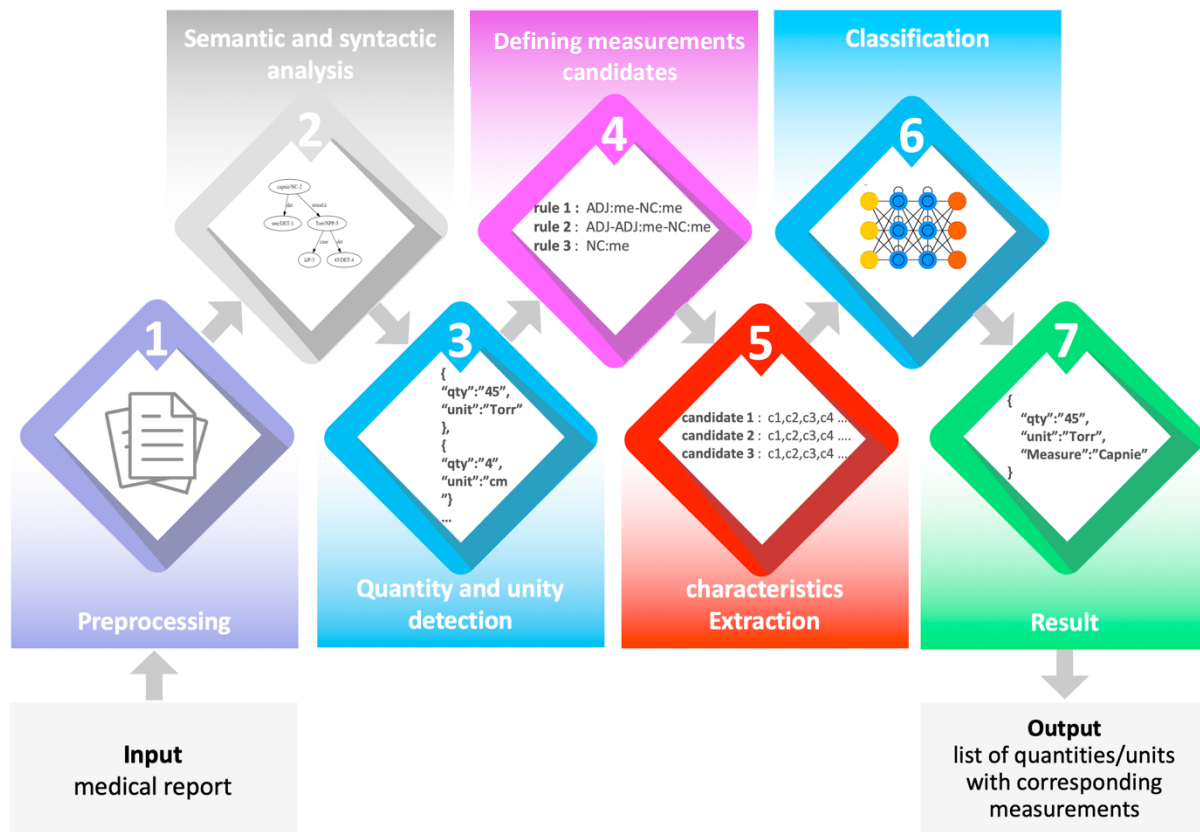


Figure 4. structure of our proposed system

3.1 Preprocessing

The input document/text is preprocessed using the same method described above (in section 2.1).

3.2 Semantic and syntactic analysis

Within this phase we apply the following tools from CoreNLP :

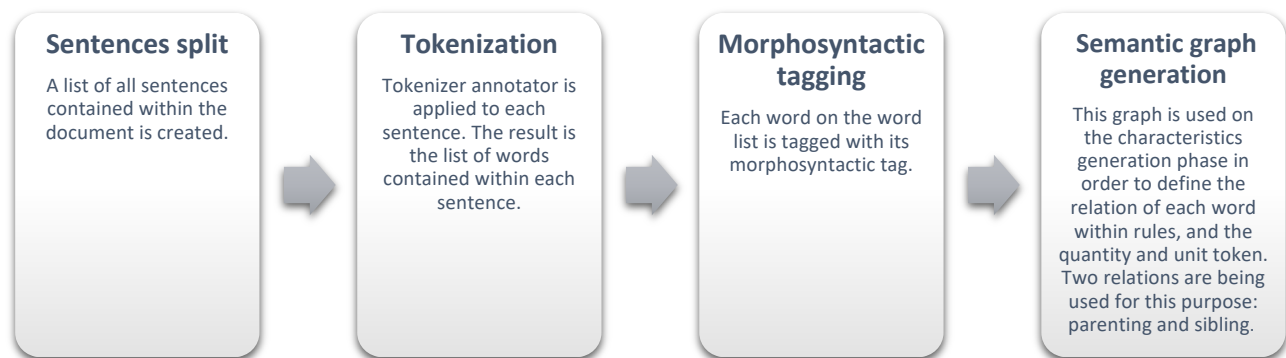


Figure 5. Semantic and syntactic analysis

3.3 Quantity/unity detection

Grobid quantities is used to detect quantities and assign units to them if they are present within the sentence. We annotate the token identified as quantity with :**q** and the unit with :**u**.

At this level, we know which tokens within the sentence are the probable quantities/units (as identified with grobid quantities) and the relationship between these token and other tokens within a given sentence.

3.4 Defining measurement candidates

For each quantity or quantity/unit couple, and based on the morphosyntactic tags of the tokens, we try to find a set of matches to the rules that was empirically defined. The result of this phase is a set of groups of words that correspond to the candidate measures found in the sentence.

3.5 Characteristics extraction

In this phase, a number of characteristics are extracted in order to be used in the next classification phase to decide which candidate measures are going to be chosen.

To do this, all tags are represented with codes (numbers). It is a kind of bijection that assigns each morphosyntactic tag used to an identifier. The purpose of this representation is to be able to use these values on our LSTM model.

3.6 Classification

In this stage, we use an LSTM neural network model to classify each measurement candidate, whether they are correct or incorrect measurement.

3.7 Measurement selection

Taking into account the result of the classification of each candidate measure, and for each quantity/unit pair, there are three possible cases:

- No candidate was classified as correct. In this case we render the measure with the highest number of tokens.
- Only one candidate measure is classified as correct, in this case it is rendered.
- If several candidate measures have been classified as correct, we take the measurement with the highest number of tokens and the closest to the quantity.

4. Results

4.1 Samples of system output

The Table 1 represents the some sentences as resulted from CoreNLP sentence split, their english translation and the list of detected measurements. The expected measurements are underlined with double lines, quantities with dotted single lines and quantities with single lines

French sentence (from report)	English translation	System output
Patient admis pour dyspnée en rapport avec un syndrome infectieux avec une <u>fièvre</u> retrouvée à 38.8 ° le jour de son hospitalisation	Patient admitted for dyspnoea in connection with infectious syndrome with a fever found at 38.8 - on the day of hospitalization	Measurement: fièvre retrouvée Quantity: 38.8 unit: °
Bilan biologique sanguin <u>hyperleucocytose</u> à 12700 éléments/mm3 avec anémie (hémoglobine à 9.1 g/dl)	Hyperleucocytosis blood biological balance at 12700 elements/mm3 with anemia (hemoglobin at 9.1 g/dL)	Measurement: hyperleucocytose Quantity : 12700 Unit : none Measurement: hémoglobine Quantity: 9.1 Unit: g/dl
Measurement 1: nous décidons de mettre en place une antibiothérapie par <u>ROCEPHINE</u> 2 g/jour pendant 10 jours	we decide to set up a ROCEPHINE antibiotic therapy 2 g/day for 10 days	Measurement: antibiothérapie par ROCEPHINE Quantity : 2 Unit : g/jour

Measurement 2: nous décidons de mettre en place une antibiothérapie par <u>ROCEPHINE 2 g/jour</u> pendant <u>10 jours</u>		Measurement: ROCEPHINE 2 g/jour Quantity: 10 Unit: jours
Les capacités de X sont tout à fait correctes puisque l'échelle de <u>raisonnement perceptif</u> est à <u>102</u> pour une échelle de <u>compréhension verbale</u> à <u>88</u> .	X's abilities are quite correct since the perceptual reasoning scale is 102 for a verbal comprehension scale at 88	Measurement: raisonnement Quantity : 102 Unit : none Measurement: compréhension verbale Quantity: 88 Unit : none
<u>troubles de la déglutition et fuite labiale liquidienne</u> évoluant depuis environ <u>2 mois</u> sans bilan réalisé	swallowing disorders and liquid lip leakage that has been evolving for about 2 months without a realized balance sheet	Measurement : fuite labiale liquidienne Quantity : 2 unit : mois

Table 1. System output samples

4.2 Grobid quantities models

The Table 2 represents the results of Grobid Quantities models training. These models are used in the third phase within our proposed system, to detect quantities and units.

Measure	Value model	Unit model	Quantity model
Accuracy	99.75 %	99.89 %	99.82 %
Precision	99.64 %	99.83 %	96.1 %
Recall	99.85 %	99.83 %	96.75 %
Score F1	99.75 %	99.83 %	96.42 %

Table 2. Grobid Quantities training results

Although the accuracy is almost perfect, the accuracy recall and F1 score are around 96%. In this case, accuracy is definitive, a model with the greatest accuracy is preferable since we want to recognize the maximum quantities/units, which is the key to recognize existing measures in a given report.

4.3 System results

The overall system results are the results of the LSTM classifier since it is the last layer in our system, that decide which tokens to annotate as measures.

The Table 3 represents the results of the most two performing configurations, the first one with 2 layers and the second one with three layers..

	LSTM1	LSTM2
Number of layers	2	3
Number of nodes within each layer	L1=100 , L2=100	L1=100 , L2=60 , L3=46
Accuracy	93.71 %	95.13 %
Precision	91.50 %	92.38 %
Recall	92.82 %	94.01 %
Score F1	88.61 %	89.49 %

Table 3. LSTM properties and raining results

5. Discussion

Although the results of the measurement extraction approach proposed in this work are promising we could improve them if we apply features engineering for the selection of features.

The current system could also be improved by applying rules for detecting measurements that follow quantity and unit, allowing more measurements to be detected.

Named entity recognition implementation to recognize measurements, can also be an addition to the system. The rules part could be completely altered, and the recognized entities could be considered as a set of candidate measures.

6. Conclusion and perspectives

Extracting measurements from medical reports represents an important phase towards their understanding. Thus, helping doctors or researchers to have a global view of all drugs and their dosages, test results, etc., will be useful in making decisions, and for drawing indicators about the patient's condition.

To meet this need, we proposed a method based on conditional random fields proposed by Grobid Quantites, for which we have prepared data to train a model to extract quantities/units. We have applied natural language processing methods for grammatical and syntactic annotation of text, and a measurement assignment system based on rules generated from corpus analysis. A final phase of classification is using the recurrent neural networks (Long Short Term Memory), which take into account the sequential aspect of the text from which we want to extract the measurements.

The current system could be used to recognize other relationships in medical reports. In fact one will have to change the part related to quantity detection and the part related to the applied rules, thus allowing to extend the operation of the system on other types of relationships.

The current system can also be used for detecting measurements in other types of documents, again the rule part may require a change.

7. References

- B. Elise, J. Vianney, M. Fleur, T. Frantz, et G. Natalia (2015). Automatic extraction of numerical values from unstructured data in EHRs, *Studies in Health Technology and Informatics*, vol. 210, p. 50–54.
- Comité éditorial pédagogique de l'UVMaF (2012-2011), *Le dossier médical*. Université Médicale Virtuelle Francophone.
- E. Ford, JA. Carroll, HE. Smith, D. Scott, JA. Cassell. Extracting information from the text of electronic medical records to improve case detection: a systematic review. *J Am Med Inform Assoc*. 2016;23(5):1007–1015. doi:10.1093/jamia/ocv180
- M. Hecquet (2012). *Le dossier médical, un outil de la relation de soins ? Éthique & Santé*, 9(1), 11–13. doi:10.1016/j.etiqe.2011.11.003
- M. Zubke (2017). Classification based extraction of numeric values from clinical narratives, In: *Proceedings of the Biomedical NLP Workshop associated with RANLP 2017*, pages 24–31, Varna, Bulgaria, 8 September 2017.
- P. Lopez (07-mai-2019). GROBID extension for identifying and normalizing physical quantities. [WWW Document, URL: <https://github.com/kermitt2/grobid-quantities>. [Accessed: 14-mai-2019].
- P. Resnik, M. Niv, M. Nossal, A. Kapit, R. Toren. Communication of clinically relevant information in electronic health records: a comparison between structured data and unrestricted physician language. *Perspectives in Health Information Management, CAC Proceedings*. 2008;Fall.
- S. Hochreiter et J. Schmidhuber (nov. 1997), « Long Short-Term Memory », *Neural Comput.*, vol. 9, no 8, p. 1735-1780.
- S. Mandhan (2015). Numerical Attribute Extraction from Clinical Texts. 10.13140/RG.2.1.4763.3365.