Digital Personalized Health and Medicine
L.B. Pape-Haugaard et al. (Eds.)

doi:10.3233/SHTI200430

Master Patient Index Standardization Patient Search Identification Service (PSIS) of the National Directorate of Health Information Systems (DNSIS) Argentina

Humberto F MANDIROLA BRIEUX[a, b] Alejandro LOPEZ OSORNIO[a] Martin DIAZ MAFFINI [a] Miguel AMORESE[a] and Daniel A RIZZATO LEDE [a]
[a] *Ministry of Health, Argentina*
[b] *HL7 Argentina.*

Abstract. Correct patient identification is the cornerstone for the proper implementation of electronic health records. Up to 20% of the registered patients are duplicated in most systems. Strong identification policies and robust systems can minimize such errors. In this poster we share the Ministry of Health recommendations for the Master Patient Index improvement using search algorithms.

Keywords. Master Patient Index, Electronic Health Records, Soundex, Levenshtein, Patient Search Identification Service

1. Introduction

One of the most serious problems for Master Patient Index (MPI) is duplication of data[1]. Conventional database systems generally use deterministic searches to find document numbers or patient names, leading to mistakes. When the person responsible for entering data in the MPI is unable to find the patient, she will reenter the information, thus creating a duplicate patient. Duplicated patients in the MPI poses the risk of the loss of data due to the fragmentation of information [4]. It is important to use Patient Search Identification Service (PSIS) to avoid these types of mistakes. By using algorithms for a probabilistic search with multiple data (name, surname, birthdate, id type and number, gender), it is possible to identify suitable candidate patients by a group of similar data[2]. PSIS allows us to identify potential similar candidates and provides significant advantages over conventional searches [3] keeping the MPI clean.

2. Materials and Methods

For our Model of PSIS we use a minimum set of patient data composed of six data points that are: **Surname** (first and following), **Name** (first and middle), **Gender**, **Document Type**, **Document Number** and **Birthdate**. We apply two combined algorithms, the first being phonetic (Soundex), the other comparative (Levenshtein).

PSIS makes the first search for candidates by using Soundex and the second level by the distance of Levenshtein's algorithm and that allows us to find the patients candidates files, see figure 1

The variables that consider are the relative weight V, the distance of Levenshtein and the length of the text of the candidate L. The coefficient F is calculated with the formula F = (L-D) / L and finally, the value of each variable is obtained by multiplying the weight V by the coefficient F. The sum of the 6 variables gives us the probability that the candidate is the wanted one, in this case, 73.75%. See figure 2

3. Results

In order to prove the advantages of the PSIS we compared the Traditional Database Search (TDS) with PSIS. Results showed that the PSIS we built was a substantial help. PSIS allows us to detect 350 duplications in an MPI that had 50.000 patients, compared to TDS detecting only 50 duplications.

4. Discussion

The PSIS significantly improves the discovery of duplicates in an MPI compared to that of TDS. PSIS searches identify 88% more duplicate candidates than those found in TDS.

5. Conclusion

If we make a mistake in the document number by a single digit, using conventional searches (TDS), the patients cannot be found. Using the Levenshtein distance algorithm allows us to identify that the number is very similar to the one searched often indicating a simple human error in data entry.

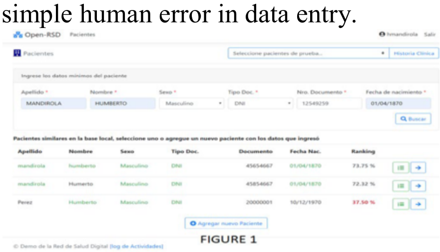
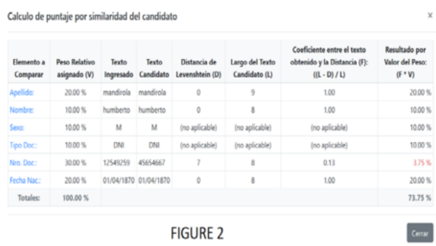


FIGURE 1



Elemento a Comparar	Peso Relativo asignado (V)	Texto Ingresado	Texto Candidato	Distancia de Levenshtein (D)	Largo del Texto Candidato (L)	Coeficiente entre el texto obtenido y la Distancia (F): ((L - D) / L)	Resultado por Valor del Peso: (F * V)
Apellido:	20.00 %	mandirola	mandirola	0	9	1.00	20.00 %
Nombre:	10.00 %	humberto	humberto	0	8	1.00	10.00 %
Sexo:	10.00 %	M	M	(no aplicable)	(no aplicable)	(no aplicable)	10.00 %
Tipo Doc.:	10.00 %	DNI	DNI	(no aplicable)	(no aplicable)	(no aplicable)	10.00 %
Nro. Doc.:	30.00 %	12549259	45654667	7	8	0.13	3.75 %
Fecha Nac.:	20.00 %	01/04/1870	01/04/1870	0	8	1.00	20.00 %
Totales:	100.00 %						73.75 %

FIGURE 2

References

[1] Just BH, Marc D, Munns M, Sandefer R. Why Patient Matching Is a Challenge: Research on Master Patient Index (MPI) Data Discrepancies in Key Identifying Fields. Perspect Health Inf Manag [Internet]. 1 de abril de 2016 [citado 23 de diciembre de 2019];13(Spring). Disponible en: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4832129/

[2] Deo RC. Machine Learning in Medicine. Circulation. 17 de noviembre de 2015;132(20):1920-30.

[3] Morris G, Afzal S, Finney D, Farnum G, Robinson C, Greene J, et al. Patient Identification and Matching Final Report. 2014;93.

[4] Patient and Sample Identification. Out of the Maze? [Internet]. [citado 24 de diciembre de 2019]. Disponible en: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5471642/