Dr. Rachael user manual

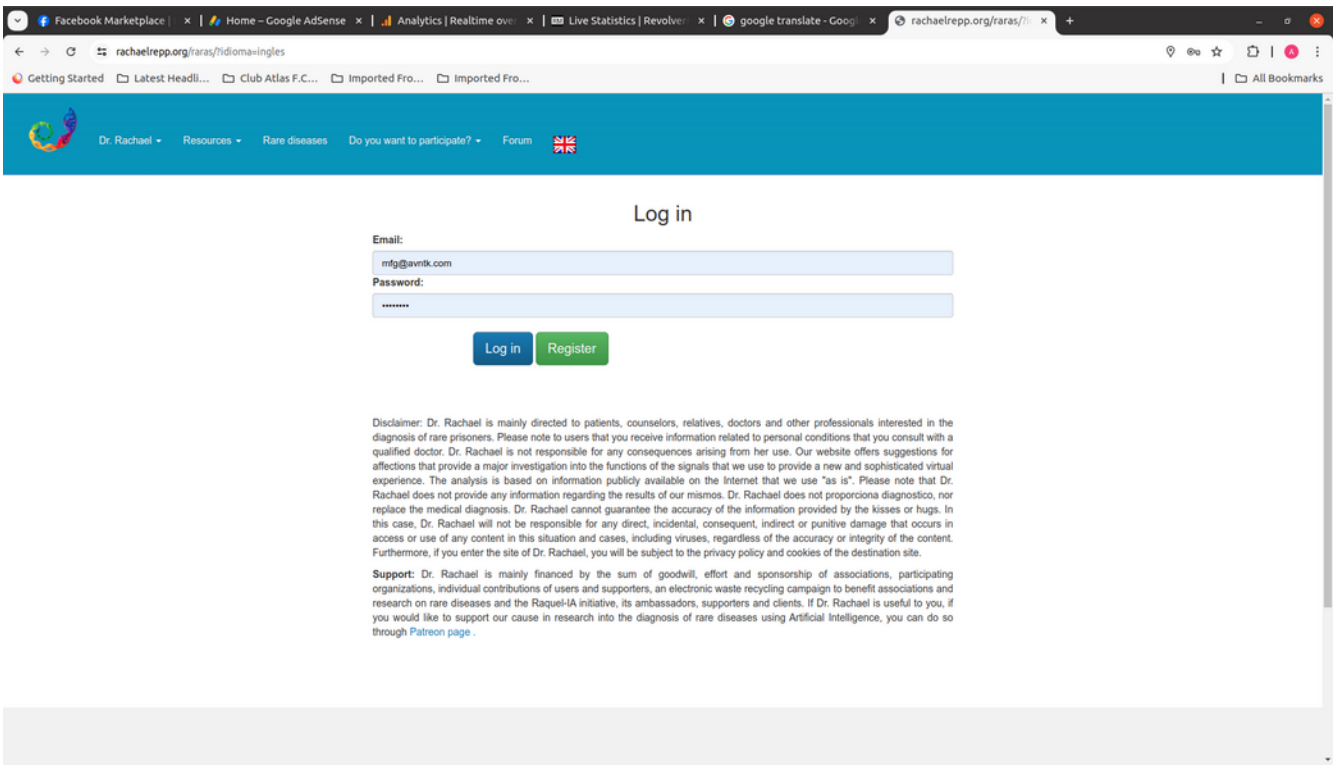
Introduction

Currently, Dr. Rachael can analyze the symptoms associated with 750 rare diseases. See the "Rare Diseases" tab for a complete list. When asked to make suggestions, it takes all conditions with a probability greater than 10% and displays them to the user, listing those with the highest probability down to the lowest probability coded in groups by color and considering the incidence of each disease. The user should click on the correct condition if known, or check “none of the above” and enter the correct diagnosis if known, before pressing “Save.”

According to a recent NORD study, more than 40% of patients with rare diseases have been misdiagnosed at least once on the path to a successful diagnosis, so validating the diagnosis by investigating other possible explanations for the identified symptoms is useful. and it will also help us validate Dr. Rachael's performance in the first instance.

Step 1

First of all, enter the website: https://rachaelrepp.org/raras/

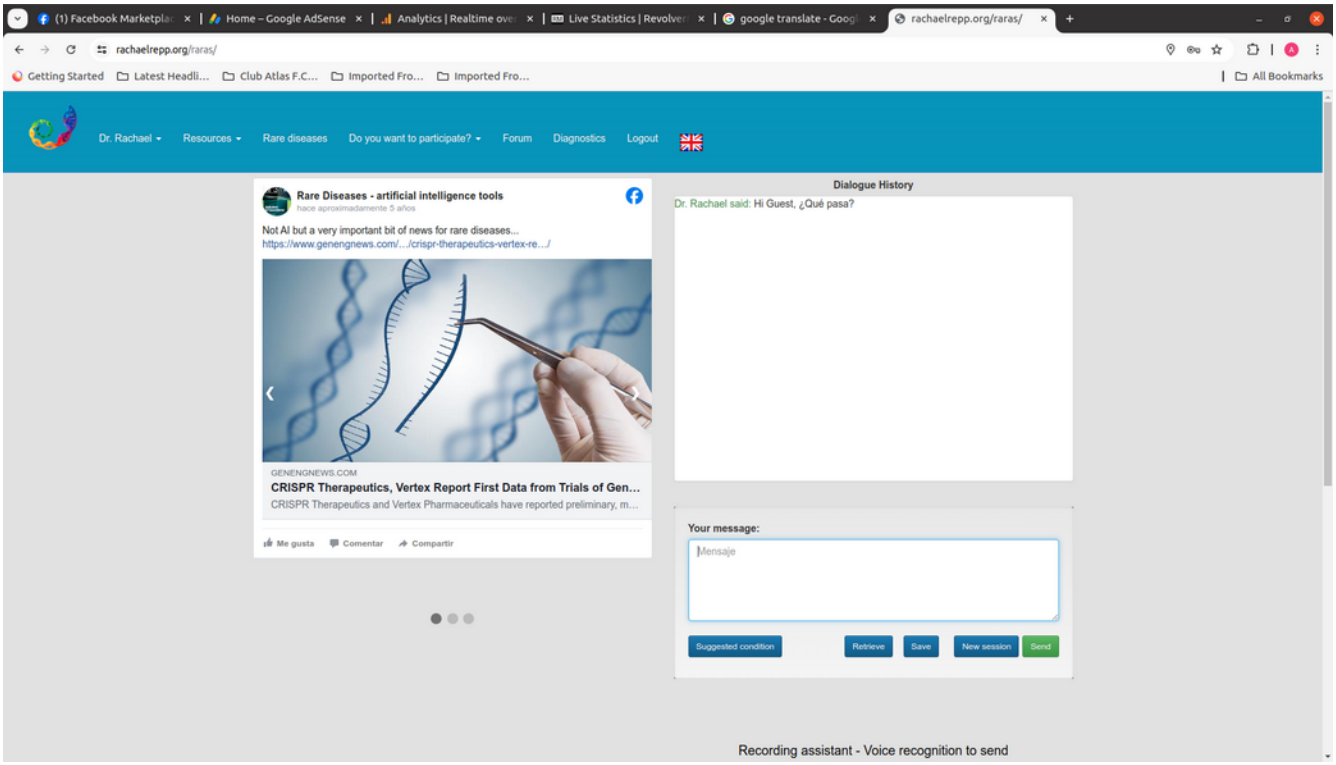


Here you can see the disclaimer that mentions that it is a research or validation site and that it is not intended to replace the diagnosis process but rather to support it. It is possible to use the system in Spanish, English, French or German. If you do not have an account, it is possible to register and use the system free of charge. For occasional visitors, we offer the following username and password:

guest@gmail.com/Marucha1 Here there is also a short description of how this project is financed and a *Patreon* link for those who are interested in supporting us through sponsorship.

Step 2

Once you enter the username and password, the main page will appear:



The knowledge base has three components: a general and personality knowledge base, a general medicine knowledge base (which needs a fair amount of additions), and a knowledge base of the most common symptoms associated with 750 of rare diseases with the highest incidence, although there are some special cases, such as diseases that are now treatable and had no cure in the past, and some diseases that were originally included for early testing purposes.

Step 3

We mainly developed the knowledge base using information from the Internet, Orphanet, NORD, NIH and GARD, so for testing purposes we suggest that you avoid these sources, but search for your own, a patient's symptoms or obtain some symptoms from other bibliographic sources. The "rare diseases" tab has a list of the rare diseases we have considered so far. These are of three types: on the left side are the diseases that Dr. Rachael can currently detect, the middle column refers to those that are currently being considered (empty), and the right side column includes conditions that we have been decided not to include for various reasons; mainly that there are a series of rare diseases that are rare but are not actually difficult to identify, others that are lethal in the short term, others are rare but benign, etc. Please note that you must enter text carefully as grammar, spelling and punctuation are important.

Descripción de síntomas

The knowledge base was generated using natural language where each symptom is described in a clear and succinct way using short phrases that describe one or two aspects in each sentence. The following are some recommendations on how to describe a patient's symptoms to make the most of Dr. Rachael's knowledge.

• Use short, clear sentences to describe each symptom, e.g. “The child has headaches often” or “The child has autism spectrum disorder.”

• Avoid giving contextual information, e.g. “The doctor told me that…” or “months later we went to a dermatologist who, upon examining her, saw…”

• Avoid detailing the specific medication that is being given to the patient, e.g. “At the age of 4 she had her first generalized seizure and since then she has been medicated… good… **first with Phenytoin and then they changed her to lamotrigine**.

• Break symptoms into separate sentences without putting things together, e.g. “she has tachycardia, a sleep disorder, and is very obsessive about buying things”… those would have to be 3 different sentences.

• The number of symptoms is important, the more symptoms you can mention, even if they seem unimportant, the more reliable the list and order of suggestions will be. Take your time and think about all the things that could be abnormal, e.g. is the head small/large, warts, freckles, are the eyes sad or almond-shaped, the ears is one higher than the other or large, what are the stools like, etc.

• It is important to mention conditions already identified: diabetes, anemia, arrhythmias, autism, epilepsy, spasms, seizures, jaundice, etc.

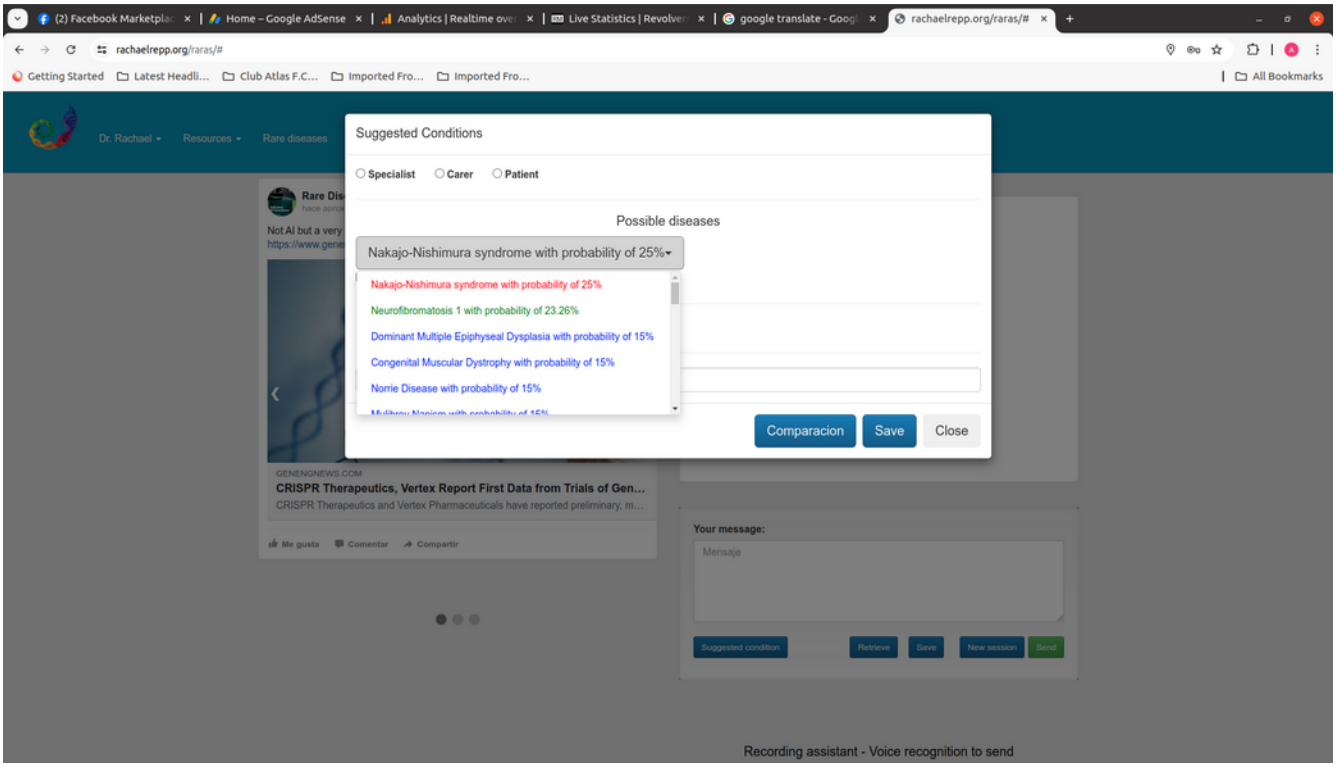
• If the system responds “I don't know how” or “undefined” it is an internal problem of the platform and the easiest thing is to send a WhatsApp to the following number to resolve it: +523339554580, Skype: gallanzi

As an example, Neurofibromatosis-1 was chosen and some typical symptoms are listed below, each of which was entered phrase by phrase into the virtual assistant.

- The child has freckles in the groin area.
- The child is short.
- The child has a large head.
- The child has bowing of the legs.
- The child has brown spots on the skin.
- The child has many benign tumors

There are a few points worth mentioning. The replies are generated by her general knowledge and her knowledge of general medicine, which are very limited, so many times the system gives replies that are not very relevant. This is something we can solve by adding general knowledge but slows the system response time. The responses have no impact on the calculations, they are only generated to make the interaction more friendly. Access to Wolfram and Wikipedia to generate replies has been temporarily restricted to make the system faster. We are working on improving the feature that allows saving sessions to make it easier for the user to add symptoms repeatedly, even if it takes a long time.

When the user clicks on “generate suggestions”, a window opens that shows the diseases that could be associated with the mentioned symptoms with more than 10% probability:



The probability calculation yields reasonable results but is not very rigorous for two reasons: the probability is taken as the number of symptoms associated with a given disease divided by the average number of symptoms in the knowledge base, corrected by an estimate of the incidence of each disease *according* to the literature. Both assumptions are questionable but give a reasonable estimate, although we still need to work on this aspect.

Another substantial challenge is that there are symptom families that are associated with only a few diseases but there are others that are associated with hundreds... that is why we have created color-coded groups of conditions that have the same probability and limited the display to 25 diseases.

Knowledge base

To save the results of a test run in the database, it is necessary to click on the correct condition if it is not in the first position and thus it will be placed in the first position, which is the one that is saved as correct. By default, the system saves the condition with the highest probability in the first position.

If the system fails and the correct disease is not listed, then the user must click “none of the above” and insert the correct condition in the space that opens.

Dr. Rachael can fail for several reasons. She may simply not have information about the disease that is related to the symptoms described; the information available in her knowledge base may need to be refined; there may be co-morbidity and some symptoms are related to another disease that the patient has leading to errors; or there may be a system error in the translation process, the ideogram (a graphical representation of a concept, in this case: symptoms) creation process, or the knowledge retrieval process.

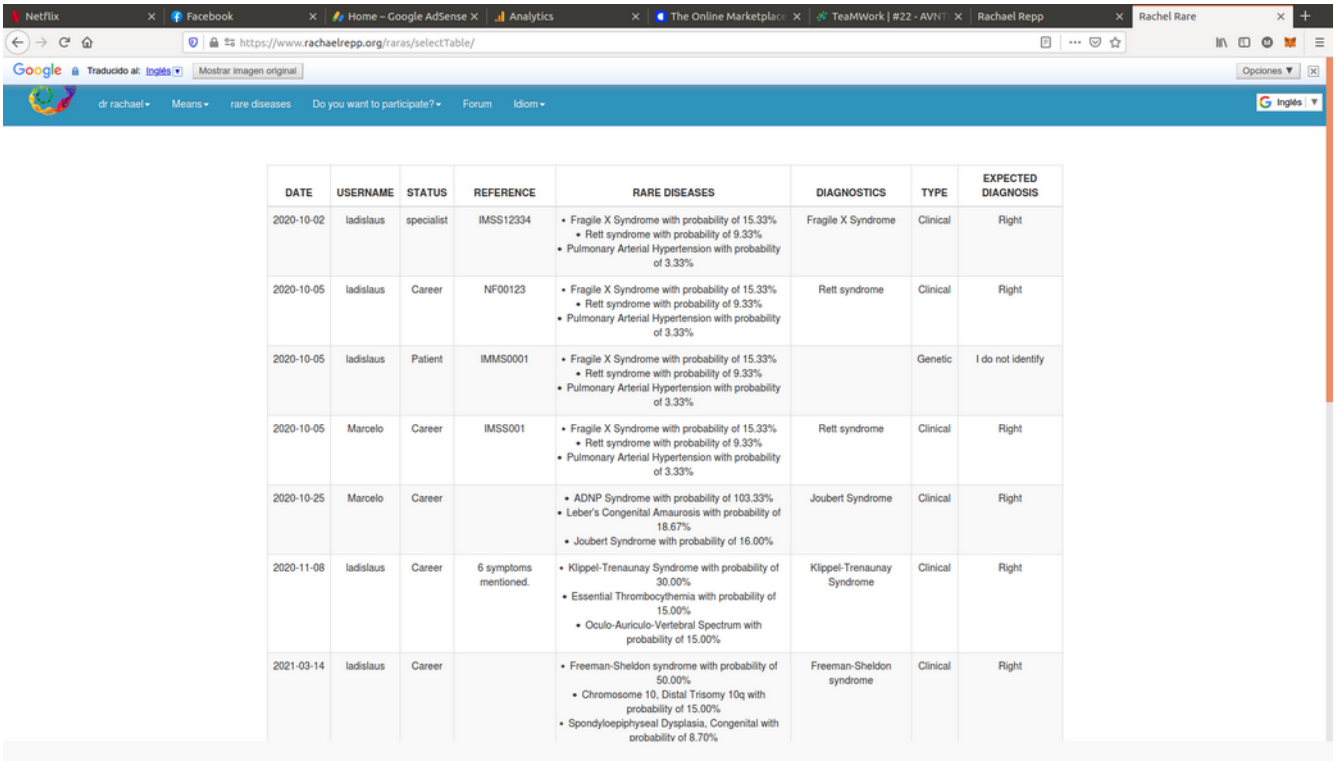
To obtain an indication of the effectiveness of the system under different conditions, we distinguish between caregiver, patient or specialist, who tend to use different language. Therefore, before pressing the “save” button, click on the corresponding buttons. Each type of user may use different vocabulary and Dr. Rachael may be sensitive to that; although it has a 35,000-word scientific dictionary available.

Furthermore, it would be useful to know, in cases where a patient's actual symptoms are described, whether the diagnosis taken as correct was provided by a genetic test or after a clinical analysis.

A “reference” text space is also provided for the user to enter a private code or identification number to allow the user to identify the patient associated with the information later, thereby respecting the patient's privacy.

Finally, the “save” button will add the result to our testing database to which we will apply some data science technique to try to help us identify errors, areas for improvement and the successful identification rate that is currently being obtained. As an indication, if only the relevant symptoms are included, the system has achieved a 75% success rate in listing the correct condition, and if at least half

of the symptoms correspond to the underlying disease, a 50% success rate has been achieved during initial tests, which were based on just 300 rare diseases. The results of each test are added and will appear in the database window which you can view by clicking "diagnostics" in the top menu:



DATE	USERNAME	STATUS	REFERENCE	RARE DISEASES	DIAGNOSTICS	TYPE	EXPECTED DIAGNOSIS
2020-10-02	ladislaus	specialist	IMSS12334	• Fragile X Syndrome with probability of 15.33% • Rett syndrome with probability of 9.33% • Pulmonary Arterial Hypertension with probability of 3.33%	Fragile X Syndrome	Clinical	Right
2020-10-05	ladislaus	Career	NF00123	• Fragile X Syndrome with probability of 15.33% • Rett syndrome with probability of 9.33% • Pulmonary Arterial Hypertension with probability of 3.33%	Rett syndrome	Clinical	Right
2020-10-05	ladislaus	Patient	IMMS0001	• Fragile X Syndrome with probability of 15.33% • Rett syndrome with probability of 9.33% • Pulmonary Arterial Hypertension with probability of 3.33%		Genetic	I do not identify
2020-10-05	Marcelo	Career	IMSS001	• Fragile X Syndrome with probability of 15.33% • Rett syndrome with probability of 9.33% • Pulmonary Arterial Hypertension with probability of 3.33%	Rett syndrome	Clinical	Right
2020-10-25	Marcelo	Career		• ADNP Syndrome with probability of 103.33% • Leber's Congenital Amaurosis with probability of 18.67% • Joubert Syndrome with probability of 16.00%	Joubert Syndrome	Clinical	Right
2020-11-08	ladislaus	Career	6 symptoms mentioned.	• Klippel-Trenaunay Syndrome with probability of 30.00% • Essential Thrombocythemia with probability of 15.00% • Oculo-Auriculo-Vertebral Spectrum with probability of 15.00%	Klippel-Trenaunay Syndrome	Clinical	Right
2021-03-14	ladislaus	Career		• Freeman-Sheldon syndrome with probability of 50.00% • Chromosome 10, Distal Trisomy 10q with probability of 15.00% • Spondyloepiphyseal Dysplasia, Congenital with probability of 8.70%	Freeman-Sheldon syndrome	Clinical	Right