

ANACPEDIA: ONLINE DICTIONARIES – AVIATION TERMINOLOGY IN PORTUGUESE, ENGLISH AND SPANISH

Fernanda Alves e Silva

National Civil Aviation Agency – ANAC/Brazil

Abstract: The National Civil Aviation Agency – ANAC/Brazil is the aviation authority responsible for aviation regulation and inspection in the country. English is the language of aviation. Brazil speaks Portuguese as the official language, what leads aviation professionals to use both English and Portuguese in their daily work, searching for translations and definitions to develop their duties appropriately. Spanish is also very used in Brazil as the country is located in South America where the majority of the states speak Spanish. A terminological resource involving these three languages is then needed and useful for different students and professionals, such as aviation personnel, translators and teachers, to name just a few. For this reason, ANAC has developed a free source of research available at <http://www2.anac.gov.br/anacpedia/>, composed of Portuguese, English and Spanish dictionaries, first presented for the ATA public during the 52nd annual conference in 2011, when ANACpedia was available for ANAC personnel only. In this paper, I intend to focus on the most recent improvements, especially in what concerns the design of a unified corpus, the use of software for concordance researches and frequency of occurrences of terms and the classification of terms in subareas. I will also indicate future intentions for the establishment of conceptual relationships between entries, selection of textual proofs, use in specialized discourse and publication of a monolingual aviation dictionary in Portuguese, an English-Spanish/Spanish-English dictionary and a list of aviation acronyms in Portuguese. Statistics of online accesses will be briefly presented.

1. INTRODUCTION

The subject of this article has to take into account the context in which it belongs, that is the Brazilian civil aviation regulation and inspection, under ANAC's responsibility. The Agency's mission and vision are to ensure all Brazilians safety and excellence in Civil Aviation and to be a reference for Civil Aviation safety and development worldwide. Our values:

- We seek to return the investment of our main shareholder: the society.
- We reject the intervention of any interest other than the public interest.
- We perform our duties with the highest level of ethics and transparency.
- We value people and meritocracy.
- We encourage innovation.
- We are proud to work at the Brazilian Civil Aviation Agency.
- Aviation is our passion.

ANAC deals with a variety of areas within aviation, the principal ones being Airport Infrastructure, Operations, Air Services, Economic Regulation of Airports, Airworthiness and International Relations.

The Linguistics Office of the Training Branch, under the Department of People Management, is responsible for ANACpedia and it develops other activities involving languages as well, such as aviation English courses and evaluations of proficiency in Spanish and English.

The Linguistics Office was presented with the challenge of developing a free, online terminology resource, which could offer aviation terms and acronyms with equivalents in other languages, definitions, synonyms and other linguistic information considering the needs of a variety of different users within the aviation area and, statistics proved later on, from outside the aviation world.

In this paper, I will focus on the most recent improvements developed by the staff in consonance with the needs of ANACpedia users and on the enrichment of information available with the use of a software for automatic extraction of linguistic and terminological data. Statistics of accesses through the Internet will be briefly presented, especially with emphasis on the variety of users from different regions of the world. In addition, I will present future intentions such as the publication of monolingual and bilingual dictionaries and a list of acronyms in Portuguese.

2. BACKGROUND

2.1 ANACpedia in 2011

ANACpedia was available on the Intranet of the Agency in 2011 and on the Internet in 2013. During the 52nd American Translators Association Annual Conference, in Boston, 2011, ANACpedia was presented and I talked about the staff involved, motivation, methodology of work, linguistic information researched, intended technical cooperation agreements, dictionaries and number of entries.

Staff was composed of language professionals and librarians. The main motivations in the beginning were to standardize translations of aviation terms from English and Spanish into Portuguese, for use in daily translation activities, and to standardize definitions. Relations between terms, presence of contexts of use and the use of tools for carrying out corpus linguistics research, for example, were not the focus. Information available at the time involved equivalent terms in Portuguese, definitions and bibliographical sources. Arrangements for a technical cooperation agreement with a Brazilian university were initiated, but not concluded. Dictionaries and entries available in 2011: English-Portuguese/Portuguese-English: 1900 terms; Spanish-Portuguese/Portuguese-Spanish: 1700 terms; Acronyms in English: 2370.

3. ANACPEDIA IN 2015

ANACpedia has been available on the Internet since 2013. The staff managing the dictionaries today is also composed of a language professional and a librarian, counting with three trainees who study Languages and/or Translation at the university. Although chiefs of sectors within ANAC have tried to expand the team, no professionals in the area of languages, linguistics or translation were hired.

A partnership with the Brazilian Department of Airspace Control started in 2014. A great part of recent researches and improvements has been developed due to the work performed by ANAC and DECEA in partnership.

The difference regarding the method of researching terms, acronyms, definitions, equivalents and other relevant information if we compare studies developed in 2011 and researches developed nowadays is that we have structured a corpus for each language involved – Portuguese, English and Spanish – and have been using them through a software for automatic extraction of information.

We can say that we have been using a methodology subsidized by corpus linguistic research emphasizing the search for relations between terms, between acronyms and between terms and acronyms, proofs of use and classification of entries in subareas of the major aviation area.

The group is prioritizing researches for enrichment of dictionaries and list of acronyms available in place of increases in the numbers of entries. Currently, the English-Portuguese/Portuguese-English dictionary has around 4000 entries; the list of aviation acronyms in English has around 4000 acronyms; and the Spanish-Portuguese/Portuguese-Spanish dictionary has around 2000 entries.

New dictionaries and a list of acronyms are also under development and the staff intends to publish them this year: an English-Spanish/Spanish-English dictionary containing around 740 terms; a Portuguese dictionary with around 2900 terms; and a list of aviation acronyms in Portuguese with around 380 entries.

4. TECHNICAL COOPERATION AGREEMENT – DETAILS

In November 2013, the team was contacted by two Translators of the Brazilian Department of Airspace Control who were highly interested in working together with the group for providing terminological data related to air space control, area in which they are specialists. DECEA Translators informed that similar initiatives were started within the Department, but interrupted due to a variety of factors.

Professionals from ANAC and DECEA noticed that a partnership for developing terminological researches and products would be very appropriate and useful. This way, the now called ANAC-DECEA working group for ANACpedia has started to meet on a regular basis since February 2014 and the parties intend to sign an official Technical Cooperation Agreement.

The aforementioned working group has been participating in very productive weekly meetings during which professionals review terminological data available, research related (RT) and new terms, discuss about technical matters as to solve doubts or to send them to specialists for resolution and elaborate and discuss about adequate classification of terms in subareas, among other things. All the information researched during meetings is gathered and recorded in terminological records for future inclusion in ANACpedia products by ANAC.

5. TERMINOLOGICAL RECORDS

ANAC-DECEA working group for ANACpedia structures the results of the weekly work in documents in *Word* format called terminological records. They aim at recording information discussed and validated during the meetings (or data to be researched in the future) in order to be included in ANACpedia by ANAC personnel, who control the software for terminology management.

Terminological records are precious sources of information as background on the discussions, validations and doubts are registered. They are preferably structured in accordance with the order in which information will be included in ANACpedia, considering the possibilities offered by the terminology management software used by the team.

6. CORPUS

ANACpedia team has structured three corpora, one for each language involved: Portuguese, English and Spanish, composed by at around 27 million words. It is important to mention that, especially for the English language, a great number of sources available on the market, such as aviation dictionaries, glossaries, lists of terms and lists of acronyms have been used, but are not included in the electronic English corpus for obvious reasons.

ANACpedia corpora is composed of a great variety of sources of research, in special publications of: the Brazilian civil aviation authority (ANAC); the Brazilian aeronautical authority; the International Civil Aviation Organization – ICAO (a specialized United Nations Agency for aviation); international aviation authorities (FAA, JAA, EUROCONTROL, JEPPESEN, Aviation Authority of Australia etc.); and the aeronautical products industry. Technical documents used in daily activities, the Internet and interactions with specialists have been largely used as sources of information.

ANAC and airspace control publications are available on the Internet, at <http://www2.anac.gov.br/biblioteca/biblioteca2.asp> and <http://publicacoes.decea.gov.br/>, with the exception of publications classified as restricted.

7. SUBAREAS

Subareas can be briefly defined as specific areas hierarchically subordinated to a major area, which, in our case, is aviation. Subareas have been researched, defined, listed, and included in ANACpedia so we can classify terms and acronyms according to their frequency in the contexts of use. The classification of entries in subareas of aviation enrich the dictionaries considerably and improves the quality of results for users, helping them to understand definitions and concepts better.

Subareas were not always present in ANACpedia. Although we had a kind of internal classification since the beginning, we had not made subareas available for users until 2014, as it is a very hard task to determine and define subareas for entry classification, especially considering aviation as a complex field. The team found that, although this is a very complex work, it was necessary to consolidate some subareas even to help the team to take decisions about synonyms and contexts, for example. An entry may have different meanings according to the subareas where they may occur. The statement of Bononno (2000, page 651) supported the decision taken by the group: “Conventional technical dictionaries, at least the better ones, often provide labels to indicate the field of application or domain for a headword.”

For a better development of our research, we have developed a list of subareas and a definition for each of them. We understand that a definition for each subarea included and prepared based on dictionaries and other sources, helps the work of classification of terms and acronyms.

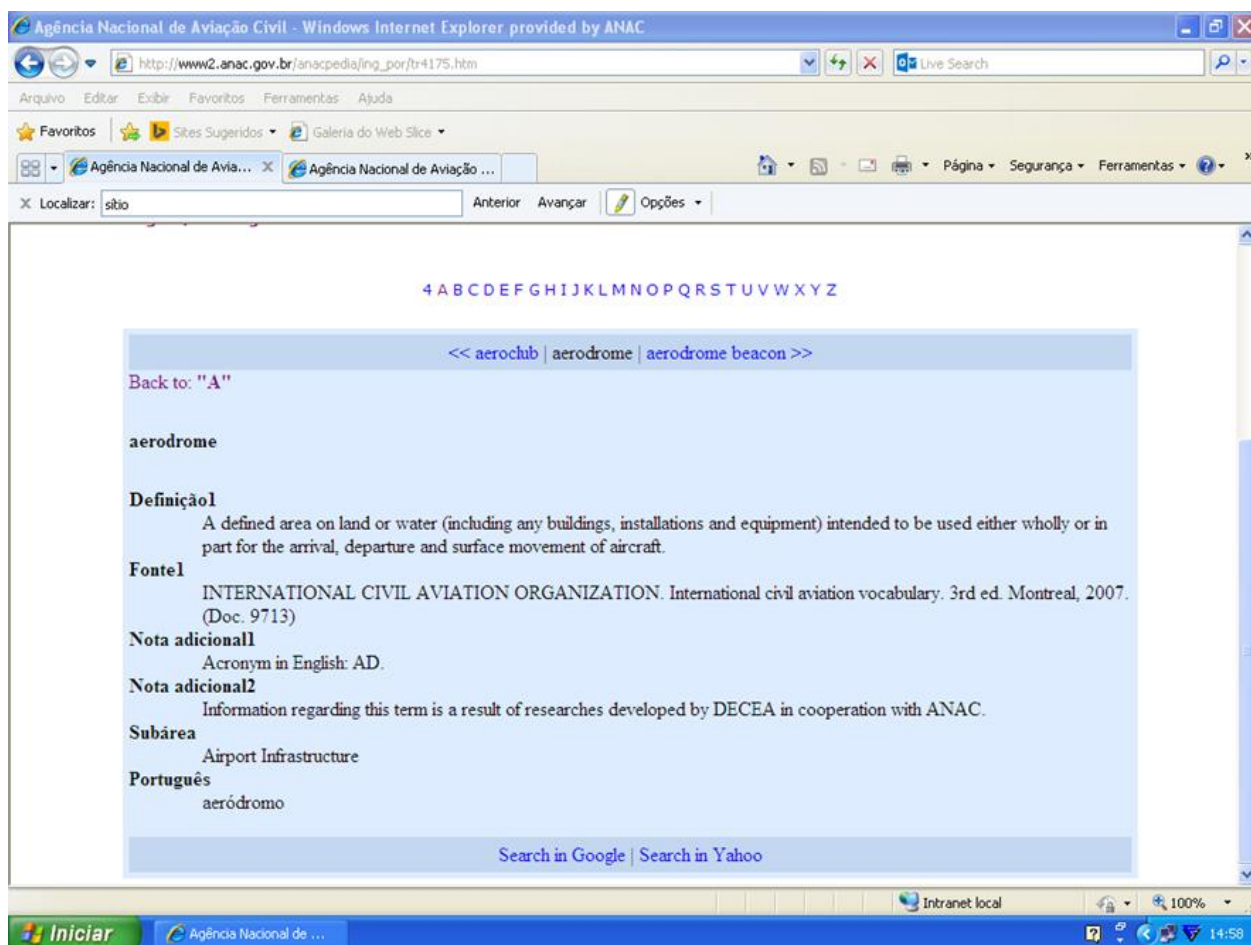
8. CORPUS LINGUISTICS RESEARCH

Since the first use of AntConc software (Concordance), it was possible to identify how useful results the software would offer. It was noticed that immediate results would open up new possibilities for researches on related, narrow (NT) and broader terms (BT), proofs of use and subareas, to name just a few.

In this item, I will detail two examples: one showing the improvements on a term resulting from the use of AntConc; and a term researched with the use of the automatic extraction tool since the beginning. Terms used as examples in English and Portuguese are *aerodrome/aeródromo* and *cockpit voice recorder (CVR)/gravador de voz na cabine de comando*.

8.1 Aerodrome/aeródromo

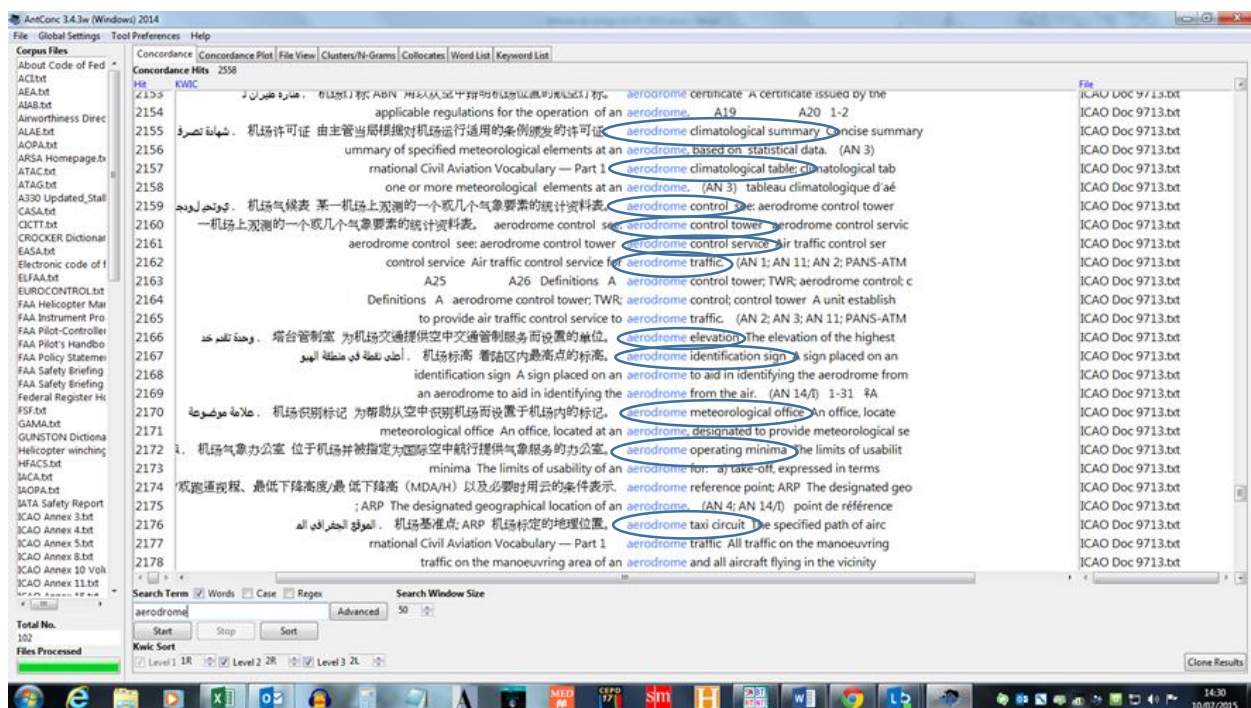
The term *aerodrome* is presented in the English-Portuguese dictionary with the following information: definition, source, two additional notes, subarea and the equivalent term in Portuguese – *aeródromo*, as can be seen below.



Picture 1 – *aerodrome*

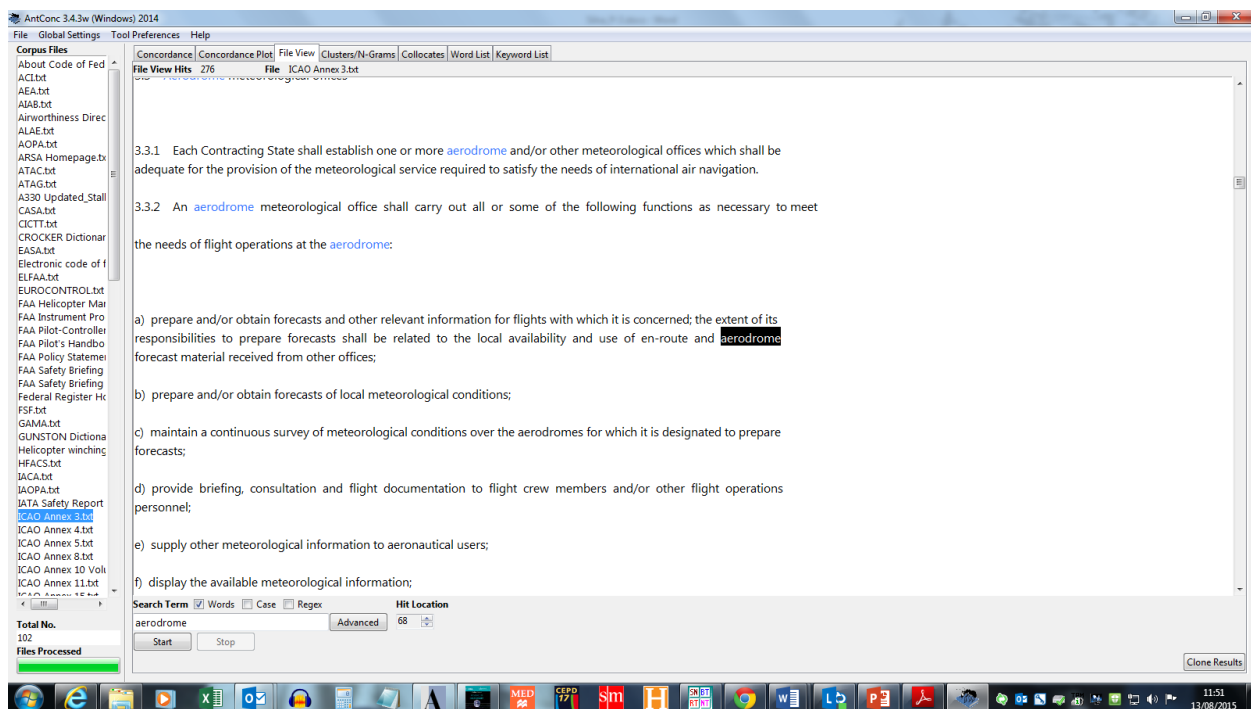
Considering *aerodrome* a term widely used in aviation, we can affirm that there is a variety of related terms linked to *aerodrome*, which are not listed in the term record available for users, as a software for automatic research was not used at the time. Neither related terms nor proofs of use in real contexts, broader terms or narrow terms, for example, were available.

If we look at picture 2, which shows part of the research using AntConc (Concordance), we notice that the software offers a list of related and narrow terms for *aerodrome*, and this considering only a small part of the results.



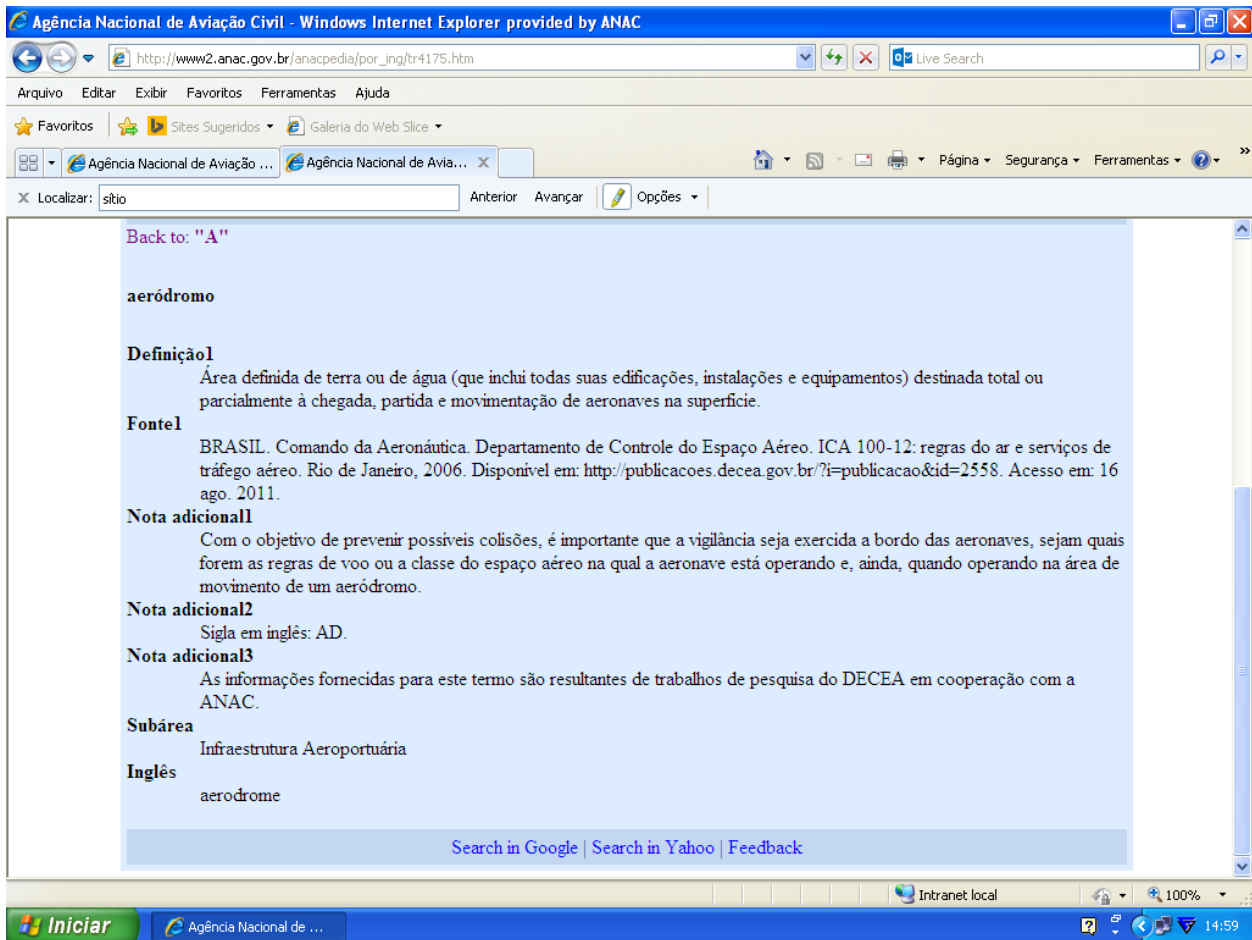
Picture 2 – AntConc (Concordance) results for *aerodrome* (RT, NT)

Picture 3 below shows proofs of use of the term *aerodrome*, in real contexts, offered by AntConc (Concordance).



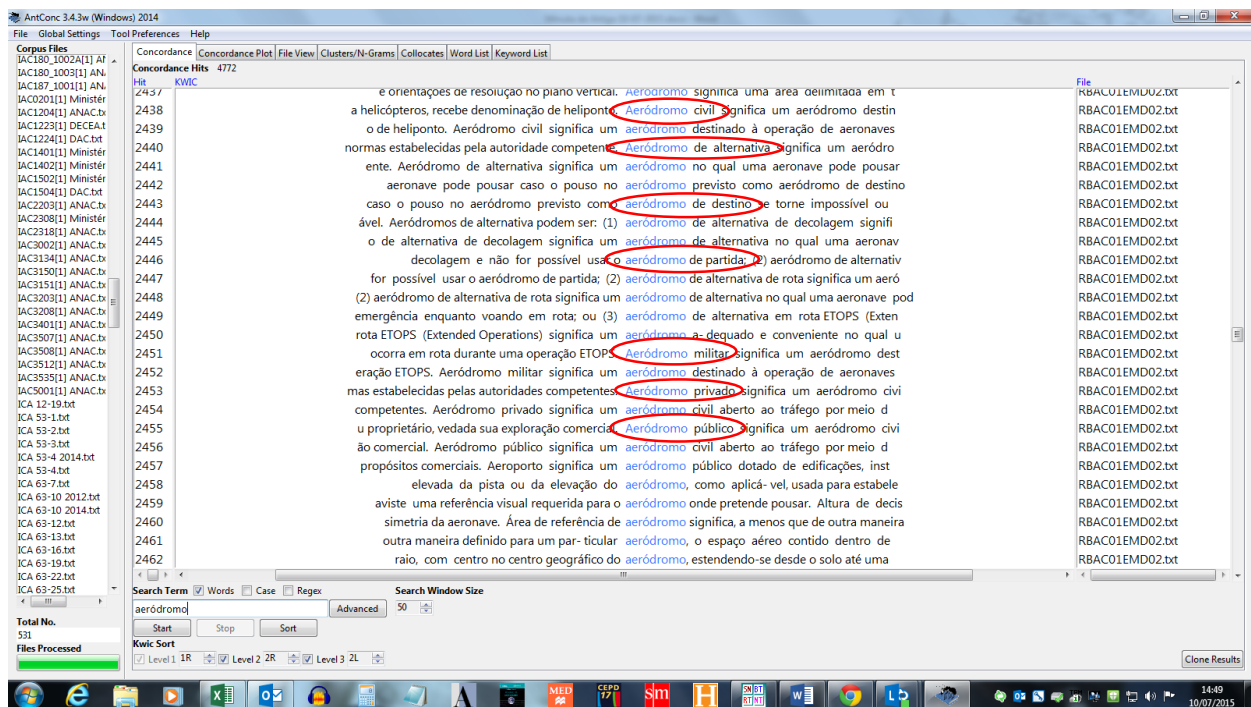
Picture 3 – AntConc (Concordance) results for *aerodrome* (proofs of use)

For *aeródromo*, equivalent in Portuguese for *aerodrome*, conclusions are the same. The original research indicated a definition, a source, three additional notes, a subarea and the equivalent term in English (*aerodrome*). However, neither related terms nor proofs of use in real contexts, broader terms or narrow terms, for example, were available. If we consider *aeródromo* one of the most used terms in aviation, the lack of information aforementioned does not enrich data available for users. The next picture shows *aeródromo* available for users before the use of an automatic tool for linguistic research.



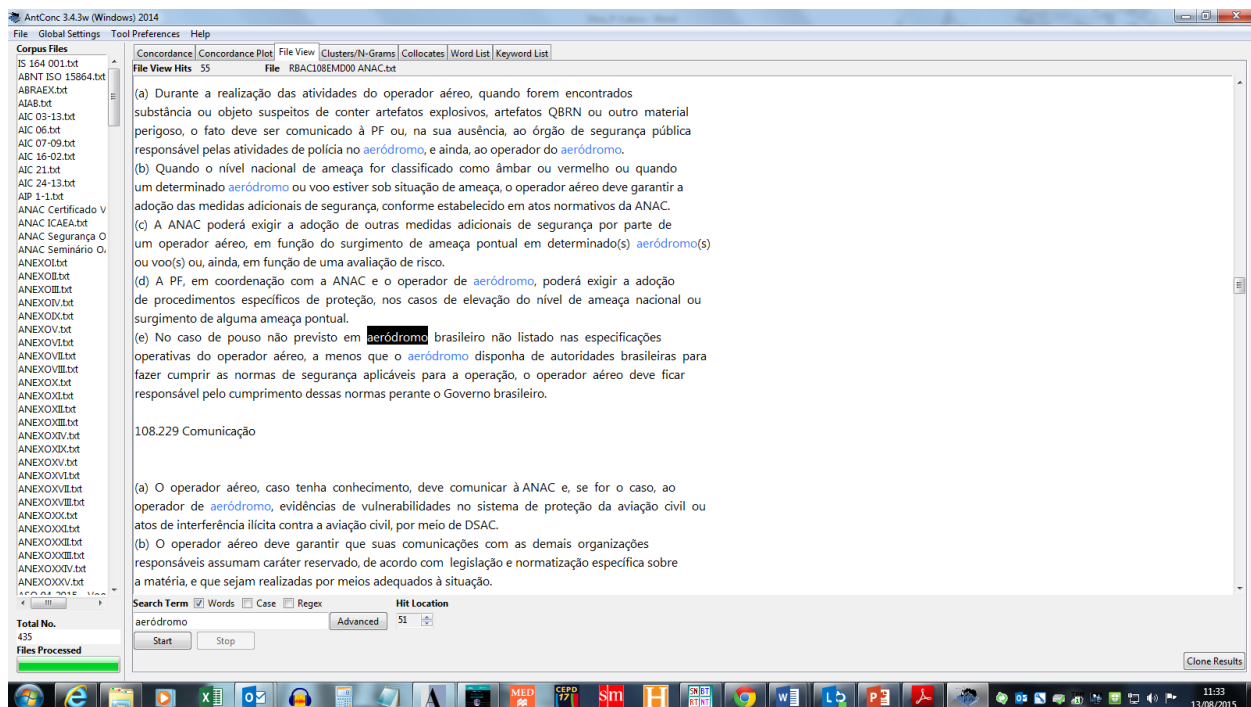
Picture 4 – *aeródromo*

The following illustration shows the possibilities of related terms and narrow terms offered by AntConc (Concordance) for *aeródromo*.



Picture 5 – AntConc (Concordance) results for *aeródromo* (RT, NT)

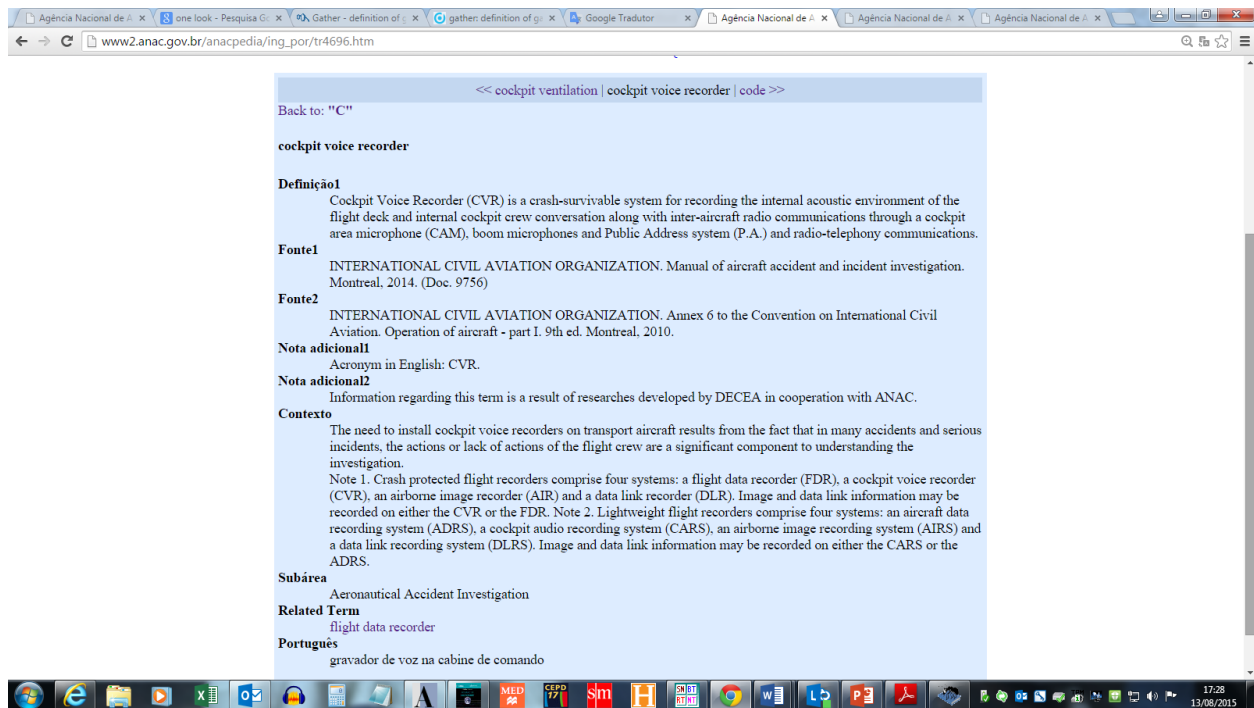
Picture 6 below shows proofs of use of the term *aeródromo*, in real contexts, offered by AntConc (Concordance).



Picture 6 – AntConc (Concordance) results for *aeródromo* (proofs of use)

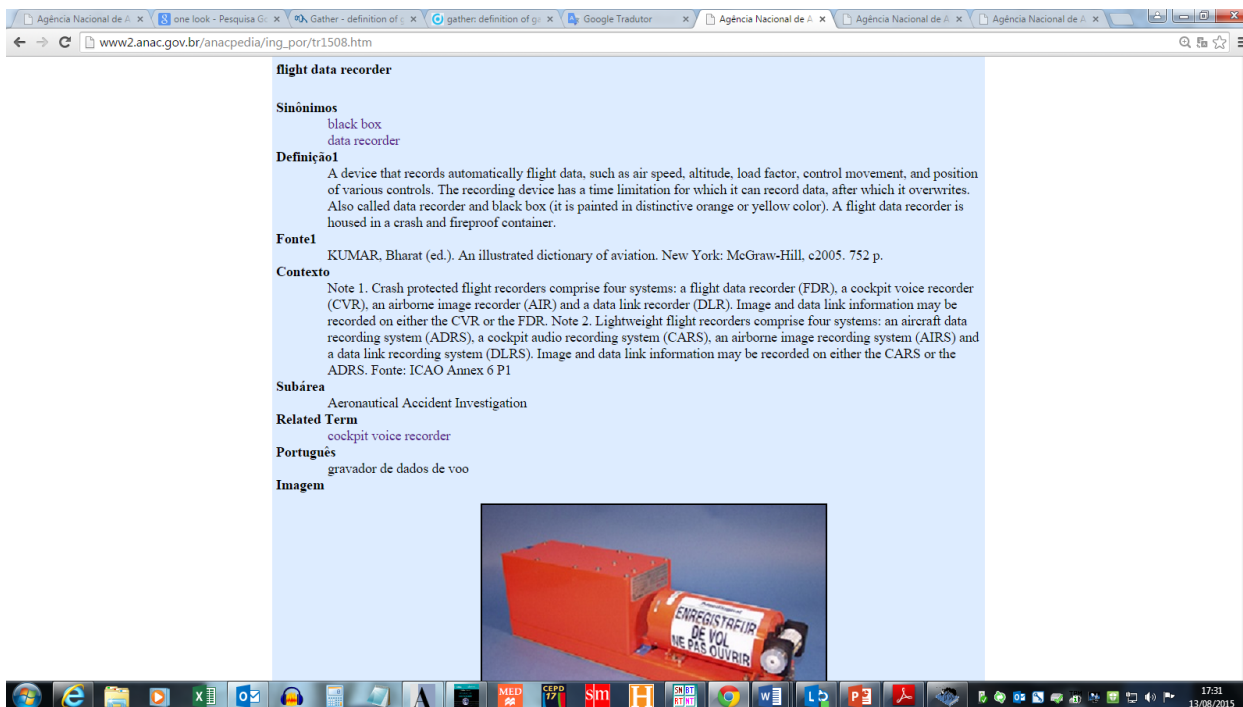
8.2 Cockpit voice recorder (CVR)/gravador de voz na cabine de comando

Cockpit voice recorder (CVR) was researched in the English and Portuguese corpora for inclusion in ANACpedia and so the automatic research made it possible the gathering of important terminological information: definition, sources, two additional notes, proof of use, subarea, related term and the equivalent in Portuguese – *gravador de voz na cabine de comando*, as below.



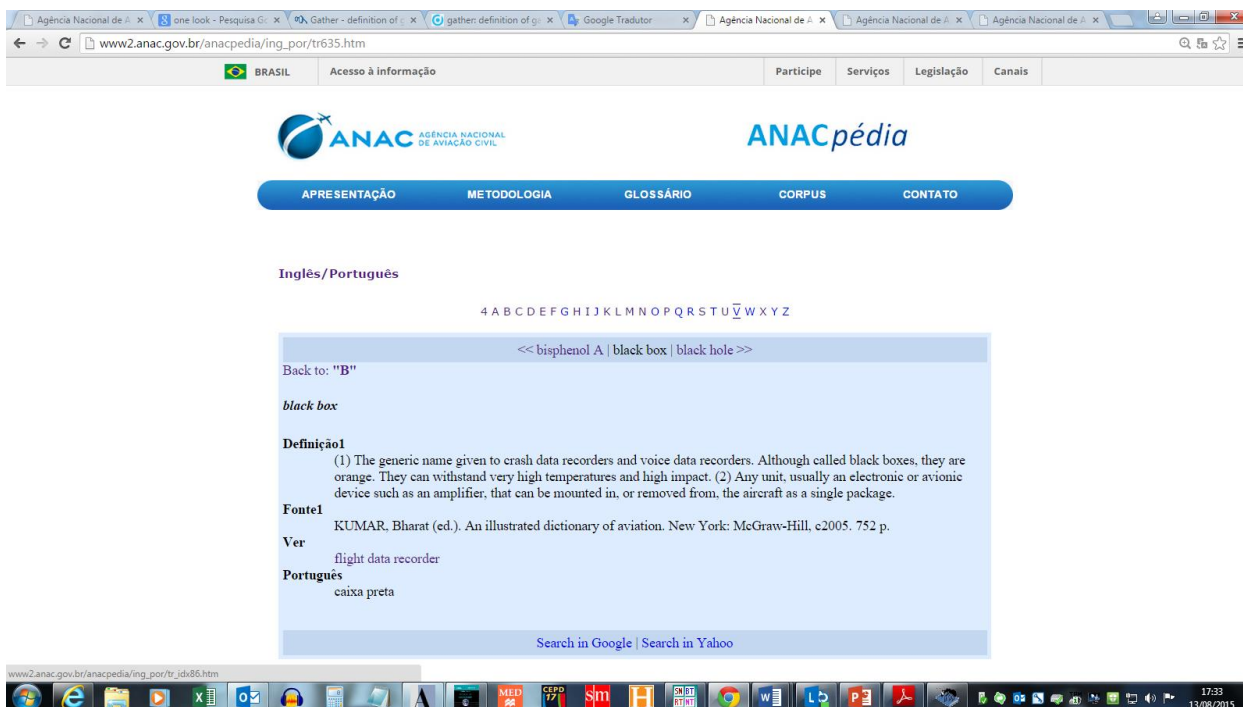
Picture 7 – AntConc (Concordance) results for *cockpit voice recorder (CVR)*

The term *flight data recorder* was found as related to *cockpit voice recorder (CVR)* in the corpora, and so the staff proceeded to researches on the RT.



Picture 8 – AntConc (Concordance) results for *flight data recorder* (RT)

When searching for information about the RT *flight data recorder*, we found two synonyms: *black box* and *data recorder*, which originated the information shown in the next two pictures.

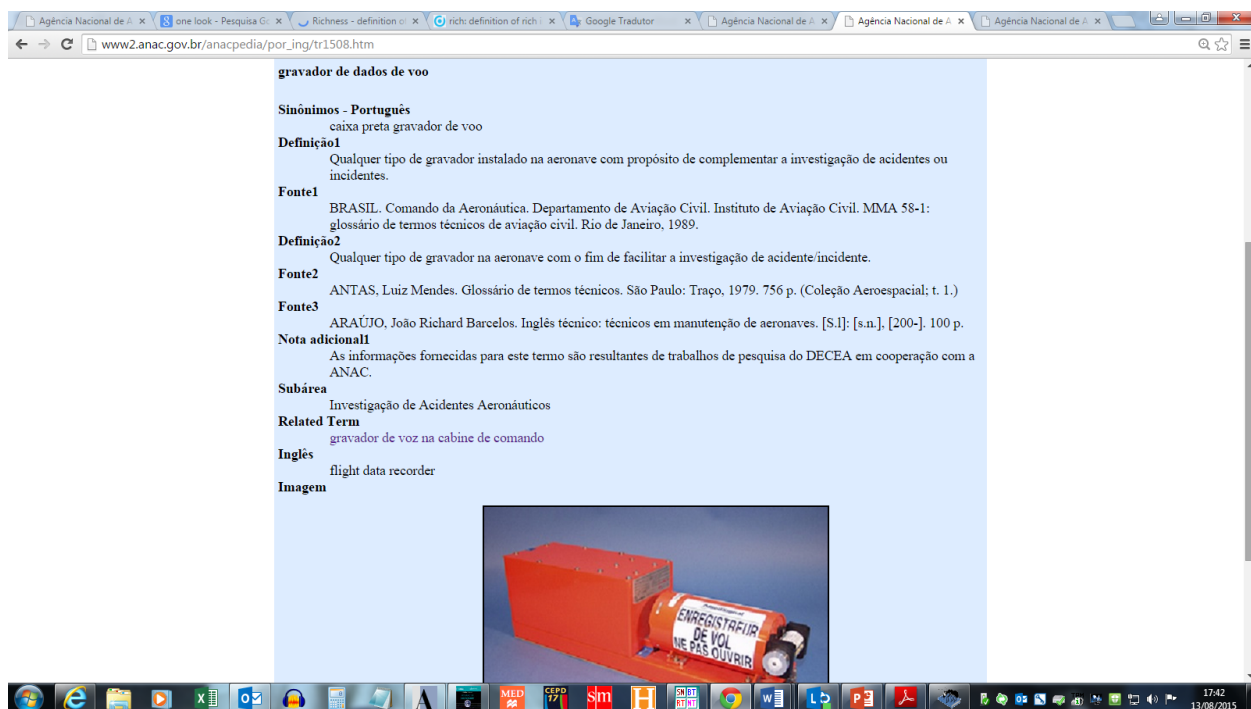


Picture 9 – AntConc (Concordance) synonym for RT *flight data recorder*



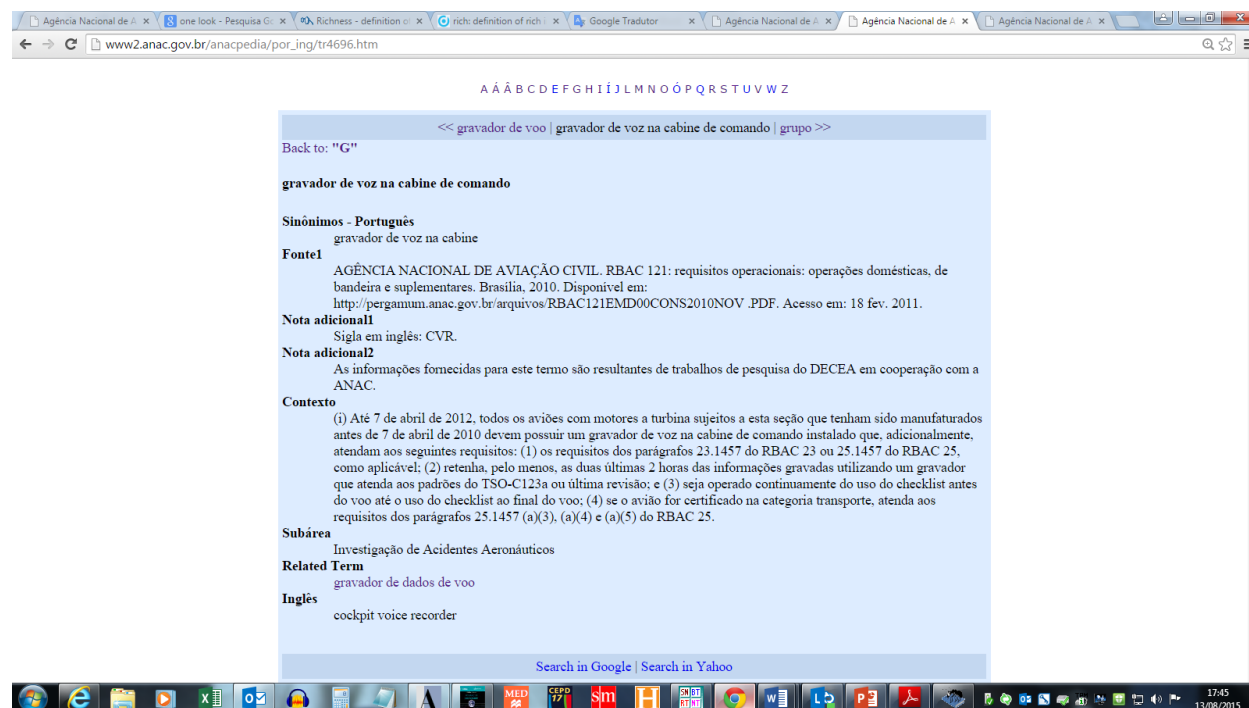
Picture 10 – AntConc (Concordance) synonym for RT *flight data recorder*

For *gravador de dados de voo*, equivalent in Portuguese for *cockpit voice recorder (CVR)*, the richness of information resulting from the search using AntConc (Concordance) can be listed as well: synonym, definitions, sources, subarea and related term.



Picture 11 – AntConc (Concordance) results for *gravador de dados de voo*

The same way that happened during researches on *cockpit voice recorder (CVR)*, the term *gravador de voz na cabine de comando* was presented in the Portuguese corpus as RT of *gravador de dados de voo*, this way leading to another semantic relation, improving information available.



Picture 12 – AntConc (Concordance) results for *gravador de voz na cabine de comando* (RT)

It can be concluded that researches with the use of a software for automatic extraction of data proved to be richer, more efficient and effective, as they do not only provide information promptly, but also lead the researcher to establish linguistic and semantic relations between terms, between acronyms and between terms and acronyms. The mentioned relations improve the quality of the terminological product in consonance with its main objective: to offer users the most comprehensible information as possible.

9. STATISTICS

Since January 2014, statistics of accesses have been monitored as we aim at verifying access rates in different periods. In addition, we would like to know the origin of accesses, that is, from which countries users have been accessing ANACpedia. Other specifications such as geographical regions in Brazil and Internet browsers used may also be monitored.

Based on the statistics available, we are positively surprised as the number of accesses surpassed the expected. The variety of countries accessing the dictionaries and list of acronyms was also a good surprise as it corroborates that ANACpedia is an useful tool for a varied audience, especially aviation professionals and translators in many countries, not only in Brazil.

Picture 13 shows the total number of accesses (13.752) from April 29 to August 12, 2015, access per page and time spent on each specific web link.

	Página	Visualizações de página	Visualizações de páginas únicas	Tempo médio na página	Entradas	Taxa de rejeição	Porcentagem de saída	Valor da página
		13.752 Porcentagem do total: 0,36% (3.819.995)	9.816 Porcentagem do total: 0,46% (2.142.763)	00:01:05 Média de visualizações: 00:01:14 (-12,37%)	2.382 Porcentagem do total: 0,20% (1.197.336)	45,09% Média de visualizações: 29,74% (51,60%)	24,28% Média de visualizações: 31,34% (-22,54%)	US\$0,00 Porcentagem do total: 0,00% (US\$0,00)
1.	/anacpedia/	5.284 (38,42%)	4.308 (43,89%)	00:00:45	842 (35,35%)	32,42%	15,86%	US\$0,00 (0,00%)
2.	/anacpedia/glosario.htm	5.009 (36,42%)	2.887 (29,41%)	00:02:01	938 (39,38%)	52,99%	36,17%	US\$0,00 (0,00%)
3.	/anacpedia/index.html	1.541 (11,21%)	1.166 (11,88%)	00:00:44	463 (19,44%)	53,56%	27,97%	US\$0,00 (0,00%)
4.	/anacpedia/metodologia.htm	1.300 (9,45%)	967 (9,85%)	00:00:16	76 (3,19%)	34,21%	9,62%	US\$0,00 (0,00%)
5.	/anacpedia/fontes.htm	158 (1,15%)	134 (1,37%)	00:00:22	14 (0,59%)	64,29%	18,99%	US\$0,00 (0,00%)
6.	/anacpedia/contatos.htm	155 (1,13%)	138 (1,41%)	00:00:43	16 (0,67%)	68,75%	27,10%	US\$0,00 (0,00%)
7.	/anacpedia/indexEnglish.html	106 (0,77%)	72 (0,73%)	00:00:29	12 (0,50%)	16,67%	11,32%	US\$0,00 (0,00%)
8.	/anacpedia/glosary.htm	79 (0,57%)	55 (0,56%)	00:00:55	6 (0,25%)	83,33%	39,24%	US\$0,00 (0,00%)

Picture 13 – number of accesses from April 29 to August 12, 2015

The next pictures show part of the Google Analytics statistical search for countries that accessed ANACpedia from March to April 2015. It is important to mention that in only one month, users from twenty different countries accessed the dictionaries and list of acronyms.

35.	/anacpedia	France	2 (0,03%)	2 (0,04%)	00:00:00	0 (0,00%)	0,00%	0,00%	US\$0,00	(0,00%)
36.	/anacpedia/	United Arab Emirates	2 (0,03%)	2 (0,04%)	00:00:27	0 (0,00%)	0,00%	0,00%	US\$0,00	(0,00%)
37.	/anacpedia/	Belgium	2 (0,03%)	2 (0,04%)	00:00:15	2 (0,21%)	0,00%	0,00%	US\$0,00	(0,00%)
38.	/anacpedia/	Canada	2 (0,03%)	2 (0,04%)	00:00:31	0 (0,00%)	0,00%	0,00%	US\$0,00	(0,00%)
39.	/anacpedia/	Chile	2 (0,03%)	2 (0,04%)	00:00:09	0 (0,00%)	0,00%	0,00%	US\$0,00	(0,00%)
40.	/anacpedia/	Japan	2 (0,03%)	2 (0,04%)	00:00:00	2 (0,21%)	100,00%	100,00%	US\$0,00	(0,00%)
41.	/anacpedia/	Singapore	2 (0,03%)	2 (0,04%)	00:00:05	0 (0,00%)	0,00%	0,00%	US\$0,00	(0,00%)
42.	/anacpedia/	Uruguay	2 (0,03%)	2 (0,04%)	00:00:16	0 (0,00%)	0,00%	0,00%	US\$0,00	(0,00%)
43.	/anacpedia/	(not set)	2 (0,03%)	2 (0,04%)	00:00:22	0 (0,00%)	0,00%	0,00%	US\$0,00	(0,00%)
44.	/anacpedia/contact.htm	Brazil	2 (0,03%)	2 (0,04%)	00:00:00	0 (0,00%)	0,00%	100,00%	US\$0,00	(0,00%)
45.	/anacpedia/fontes.htm	United States	2 (0,03%)	2 (0,04%)	00:00:07	2 (0,21%)	0,00%	0,00%	US\$0,00	(0,00%)
46.	/anacpedia/glossario.htm	France	2 (0,03%)	2 (0,04%)	00:00:28	0 (0,00%)	0,00%	0,00%	US\$0,00	(0,00%)

Picture 14 – countries that accessed ANACpedia from March to April 2015

47.	/anacpedia/glossario.htm	India	2 (0,03%)	2 (0,04%)	00:00:00	2 (0,21%)	100,00%	100,00%	US\$0,00	(0,00%)
48.	/anacpedia/glossario.htm	Italy	2 (0,03%)	2 (0,04%)	00:00:00	2 (0,21%)	100,00%	100,00%	US\$0,00	(0,00%)
49.	/anacpedia/glossario.htm	(not set)	2 (0,03%)	2 (0,04%)	00:00:00	0 (0,00%)	0,00%	100,00%	US\$0,00	(0,00%)
50.	/anacpedia/index.html	Argentina	2 (0,03%)	2 (0,04%)	00:00:00	0 (0,00%)	0,00%	100,00%	US\$0,00	(0,00%)
51.	/anacpedia/indexEnglish.html	Portugal	2 (0,03%)	2 (0,04%)	00:00:24	0 (0,00%)	0,00%	0,00%	US\$0,00	(0,00%)
52.	/anacpedia/metodologia.htm	Cape Verde	2 (0,03%)	2 (0,04%)	00:00:03	0 (0,00%)	0,00%	0,00%	US\$0,00	(0,00%)
53.	/anacpedia/metodologia.htm	France	2 (0,03%)	2 (0,04%)	00:00:00	0 (0,00%)	0,00%	100,00%	US\$0,00	(0,00%)
54.	/anacpedia/metodologia.htm	Kenya	2 (0,03%)	2 (0,04%)	00:00:00	2 (0,21%)	100,00%	100,00%	US\$0,00	(0,00%)
55.	/anacpedia/metodologia.htm	United States	2 (0,03%)	2 (0,04%)	00:00:03	0 (0,00%)	0,00%	0,00%	US\$0,00	(0,00%)
56.	/anacpedia/metodologiaEsp.html	Brazil	2 (0,03%)	2 (0,04%)	00:00:22	0 (0,00%)	0,00%	0,00%	US\$0,00	(0,00%)

Linhas 1 - 56 de 56

Picture 15 – countries that accessed ANACpedia from March to April 2015

10. FINAL REMARKS

ANACpedia allows endless possibilities for the development of actions, activities, works, new dictionaries and databases, ontologies, classification systems, domain trees, etc., not to mention academic studies on the project.

Therefore, given the numerous possibilities, it is important to plan actions in order to set priorities, while considering results in the medium and long term.

As next steps, the team intends to develop a monolingual dictionary in Portuguese, a list of aviation acronyms in Portuguese, and two bilingual dictionaries: an English-Spanish/Spanish-English and a French-Portuguese/Portuguese-French dictionary. In addition, continuous review and improvement of dictionaries and list of acronyms available is constant in our daily activities for quality assurance. The celebration of Technical Cooperation Agreements is also desired.

11. REFERENCES

1. ANAC's Training Web Page <https://sistemas.anac.gov.br/capacitacao/> in August 6th, 2013.
2. ANAC's Web Page <http://www2.anac.gov.br/portal/cgi/cgilua.exe/sys/start.htm?sid=330> in August 5th, 2013.
3. ANTHONY, Laurence. **AntConc (Windows, Macintosh OS X, and Linux)**: build 3.4.3. Tokyo: Waseda University, september, 17th, 2014. Available in: <http://www.laurenceanthony.net/software/antconc/releases/AntConc343/help.pdf>
Access on: July 24th, 2015.
4. BONONNO, R. Terminology for translators: an implementation of ISO 12620. Meta, XLV, 4, 2000.
5. BRASIL. Agência Nacional de Aviação Civil. ANACpédia. Available in: <http://www2.anac.gov.br/anacpedia>.