**Introduction**

The *Similarity Check* service has been developed to allow editors to check the similarity of submitted manuscripts with already published scientific articles and web pages. For this check, a text comparison tool called *iThenticate* is used*.*

The basic objective of the service is to help researchers avoid accidental plagiarism of others' or their own previously published work.

The process is based on the comparison of a manuscript and a huge full-text database *(Turnitin)* of existing scientific articles. The *Turnitin* database is constantly growing, always up to date, and based on the following items: one billion web pages, 57 million pieces of content from [*Crossref*](https://www.crossref.org/services/similarity-check/) (they provide the *DOI* for the articles), 100 million content items from other content providers such as [*Pearson*](https://www.pearson.com/)*,* [*Cengage,*](https://www.cengage.uk/)[*EBSCOHost, etc.*](https://www.ebsco.com/products/ebscohost-research-platform)

**How does the service work?**

After uploading the document to be examined, a so-called similarity report is generated, which you can compare with the results page by page. In this way, you can find whether the found similarity is signed or it requires further examinations.

During this examination, it is possible to discover where the similarity is originated, and the original document can also be read by clicking on the marker.

It also shows where the quote comes from.

Matching text does not necessarily mean plagiarism: it can be correctly quoted, or it can use for example, standard or general scientific terms.

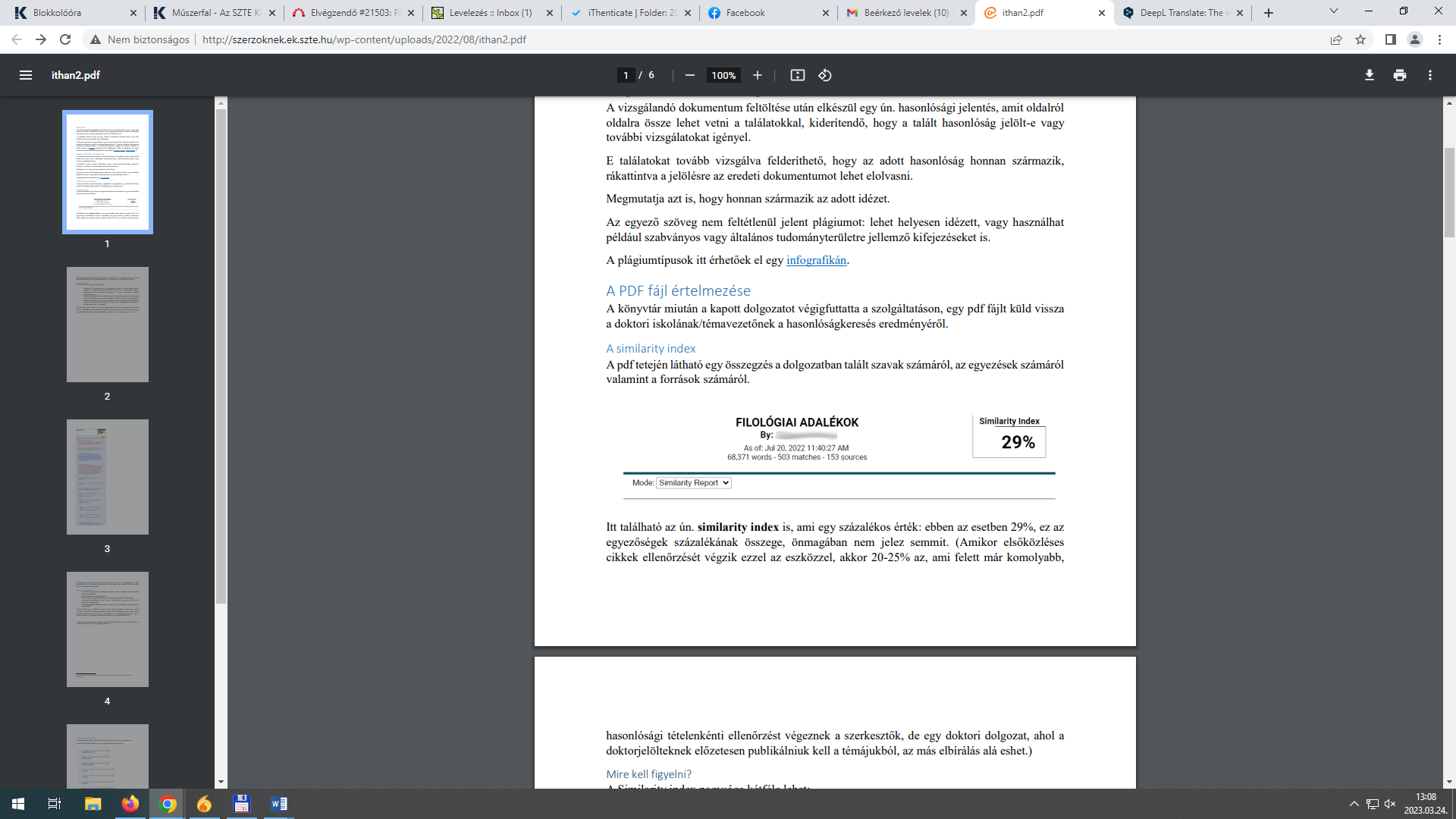
The types of plagiarism are available in an infographic [HERE.](https://www.ithenticate.com/resources/infographics/types-of-plagiarism-research)

**Interpreting the PDF file**

The library checks the received document with the service, then returns a pdf and a html file to the doctoral school which contains the results of the similarity search.

***The similarity index***

At the top of the pdf, you can see a summary of the number of words found in the paper, the number of matches, and also the number of sources.



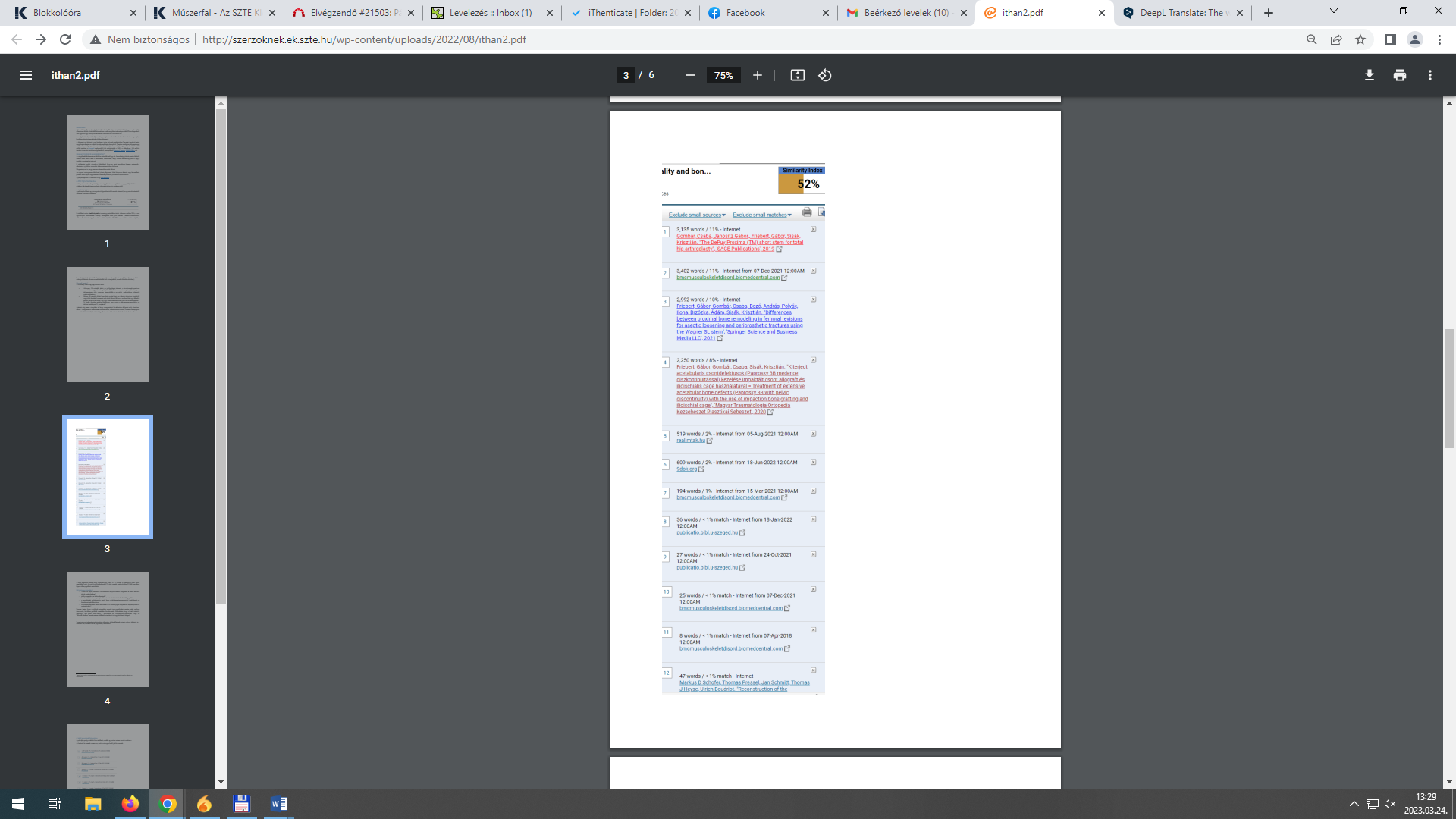
Here you will also find the so-called **similarity index,** which is a percentage value: in this case it is 29%, this is the is the sum of the percentages of similarities, and it does not indicate anything in itself. (When first-read articles are checked with this tool, 20-25% is considered a more serious similarity, then editors usually check each single similarity item. However, PhD-candidates have to public articles in their field before submitting their thesis, so their work are assessed in a different way. We also note that undergraduate theses are assessed less strictly.)

***What should you consider?***

There are two types of the size of the *Similarity index*:

* Low (below 20 percent) value also deserves attention. In that case, the following issues should be checked: are the references adequate and the matching texts are marked? You should also check whether the document is closely enough related to the literature available in the field.
* If the value of the similarity index is high (above 20 percent,) there may be either one significant quote from a single source or many short quotes from several sources. The document may contain a quotation of a single primary source, or – if it is a synthesis work – it can use multiple sources. In this case, it must be checked whether the document meets the point No. 21 of the regulations related to the PhD-theses.

It is also recommended to check which part of the thesis contains many references. It can be acceptable in the review of the literature, in the methodological section, and also in the description of the materials and tools. On the other hand, it is less acceptable in the conclusion part of the thesis.



In the picture above, you can see that the similarity index is 52%, but most of them originate from the author’s own works. The following values are below 1%, these are negligible compared to the 3268080 words of the whole thesis.

***What should you consider?***

* the level of acceptance of the use of previous own publications in the regulation of the specified PhD--program?
* are self-references indicated?
* to what extent does the journal contribute to the republishing of excerpts? See an example.[[1]](#footnote-1)
* whether the co-authors have stated how much the doctoral student has contributed to the cited publication
* whether the co-authors and copyright holders of previously published articles have permitted the republishing?

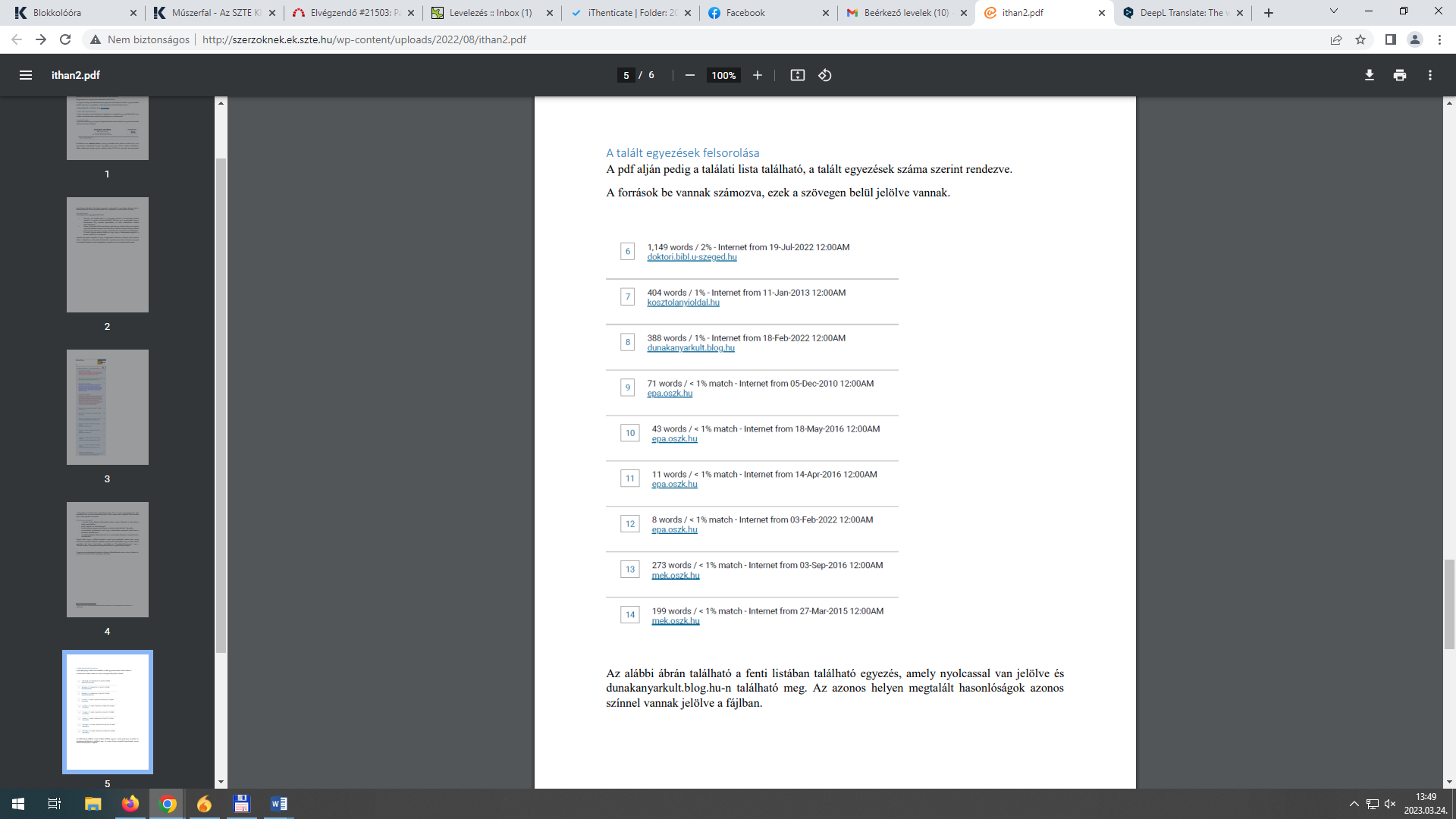
It is very important that candidates respect copyright rules when using their own, possibly co-authored, previously published works. A written permission of the publisher may be required from the candidate. Many publishers’ website contains instructions and/or forms to fill in under *“Permissions”* or *“Authors”* menus.

Of course, primary text citations may also appear, depending on the specified field: literary work, historical sources, or legislations can be quoted.

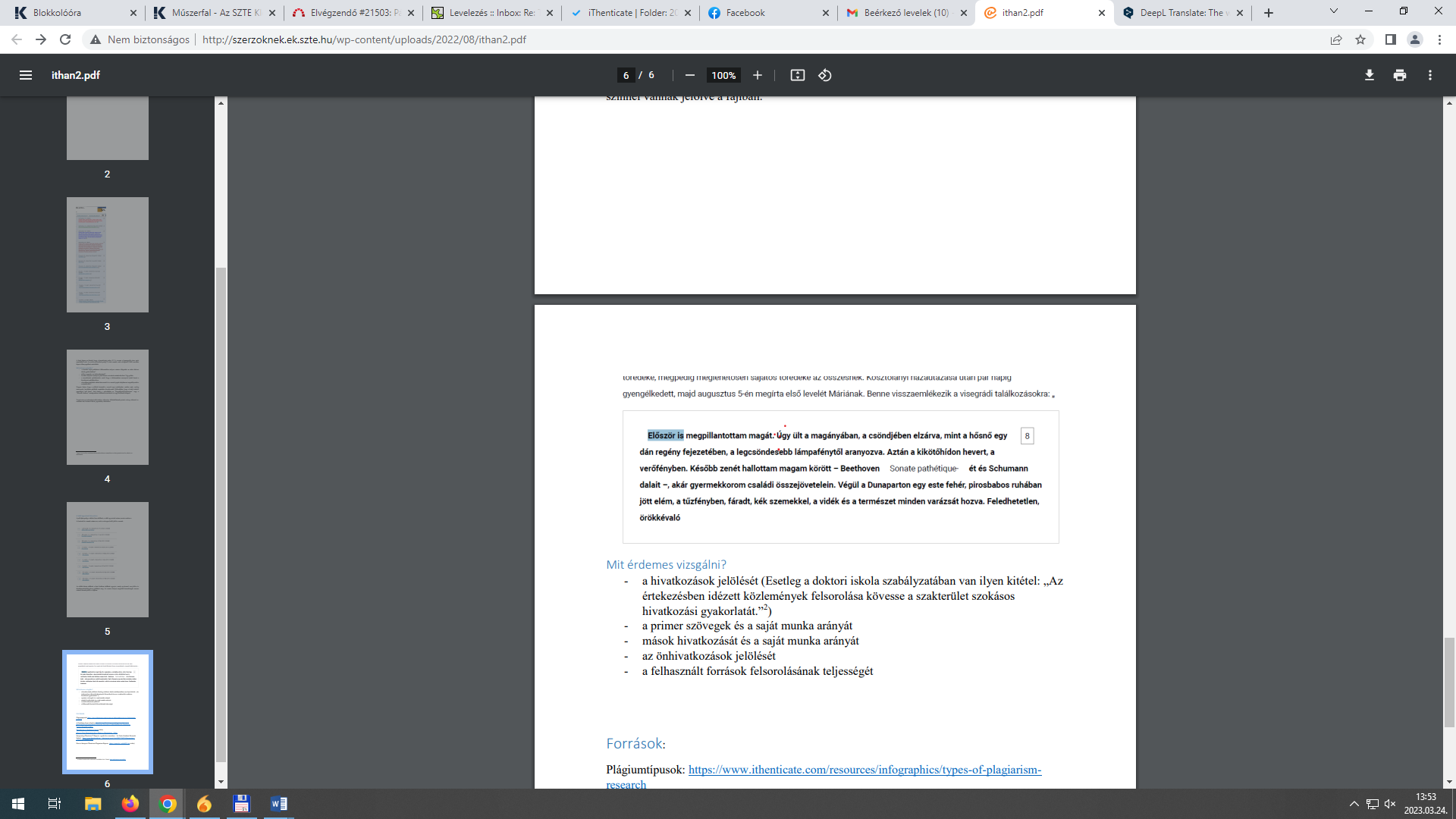
***The list of matches (similarities) found***

At the bottom of the pdf, you can find the list of matches, sorted by the number of matches found.

The sources are numbered and are marked within the text.



The figure below shows the match in the list above, marked with an eight and found on *dunakanyarkult.blog.hu.* Similarities found in the same place are marked in the same color in the file.



***What should you consider?***

* the marking of references (Maybe the regulations of the doctoral school contain a clause:: “The list of publications cited in the thesis must follow the usual citation practice in the field.”[[2]](#footnote-2)
* the ratio of primary texts and own work
* the marking of self—references
* the whole list of sources used in the thesis

**Sources of this information sheet:**

* Types of plagiarism: <https://www.ithenticate.com/resources/infographics/types-of-plagiarism-research>
* The meaning of the *Similiraty Report:* <https://www.ithenticate.com/plagiarism-detection-blog/bid/63534/CrossCheck-Plagiarism-Screening-Understanding-the-Similarity-Score#.WJuaXYVOK71>
* [Introduction to Similarity Check](https://www.youtube.com/watch?v=O-CuUA2Mzq4) (video)
* [How to Clear Permissions for a Thesis or Dissertation - Wiley](https://www.wiley.com/en-us/network/publishing/research-publishing/trending-stories/how-to-clear-permissions-for-a-thesis-or-dissertation)
* Interpreting iThenticate™ Reports: a guide for researchers – La Trobe Graduate Research School - <https://www.latrobe.edu.au/__data/assets/word_doc/0003/784236/InterpretingiThenticate-Reports.docx>
* How to Interpret iThenticate Plagiarism Reports - <https://www.youtube.com/watch?v=_ns9srRX_kghttps://youtu.be/_ns9srRX_kg%20> (videó)

1. https://www.wiley.com/en-us/network/publishing/research-publishing/trending-stories/how-to-clear-permissions-for-a-thesis-or-dissertation [↑](#footnote-ref-1)
2. Principles for editing and reviewing doctoral theses. Source: <http://phd.szote.u-szeged.hu/index_en.htm> [↑](#footnote-ref-2)