

Porozmawiajmy o wspomaganiu komputerowym w pracy tłumacza

Krzysztof Matecki

Wielkopolskie Koło PT TEPIS

O czym będzie prezentacja

1. Miejsce, z którego można pobrać pliki z archiwami pamięci tłumaczeniowych DGT-TM.
2. Omówienie zawartości archiwów DGT-TM.
3. Scalanie plików w pobranych archiwach DGT-TM do plików zawierających dane bilingwalne - pary językowe.
4. Edycja danych w scalonych archiwach DGT-TM na tekst niesformatowany.
5. Kompilowanie danych w oprogramowaniu XAIRA w celu stworzenia korpusu bilingwalnego.

DGT-Translation Memory - EU Science Hub

Dyrekcja Generalna ds. Tłumaczeń
Pisemnych Komisji Europejskiej
udostępnia od 2007 roku
wielojęzyczną pamięć
tłumaczeniową (DGT-TM),
wspierając wysiłki Komisji
Europejskiej na rzecz
wielojęzyczności, różnorodności
językowej, jak również ponownego
wykorzystania informacji
pochodzących od Komisji.



The screenshot shows the EU Science Hub website. The header includes the European Commission logo and the text "EU SCIENCE HUB" and "The European Commission's science and knowledge service". The breadcrumb trail is "European Commission > EU Science Hub > Language Technology Resources > Dgt translation memory". The navigation menu includes "Home", "About Us", "Research", "Knowledge", "Working With Us", "Procurement", and "News & Events". The main content area is titled "DGT-Translation Memory" and lists several resources. The "DGT Translation Memory" link in the left sidebar and the "Download the DGT Translation Memory" link in the main list are highlighted with red boxes.

Language Technology Resources

- JRC-Acquis: a multilingual parallel corpus
- DGT-Acquis: multilingual parallel corpora from the EU Official Journal
- DCEP: Digital Corpus of the European Parliament
- DGT Translation Memory**
- ECDC Translation Memory
- EAC Translation Memory
- JRC-Names: names and their spelling variants
- JRC Eurovoc Indexer

DGT-Translation Memory

- Introduction
- DGT's Translation Memory
- Description of the Data - Pre-processing
- Statistics for the DGT Translation Memory
- Conditions for Use
- Difference between the DGT Translation Memory and the other resources available here
- **Download the DGT Translation Memory**
- How to produce bilingual extractions
- More details / Reference publication
- Acknowledgement and Contact
- ISLRN: 10-653-952-884-4.

<https://ec.europa.eu/jrc/en/language-technologies/dgt-translation-memory>

Pobieranie archiwów z plikami

Home About Us Research Knowledge Working With Us Procurement News & Events Our Communities

Strona EU Science Hub,
z której można pobrać
archiwa z pamięcią
tłumaczeniową DGT-TM.

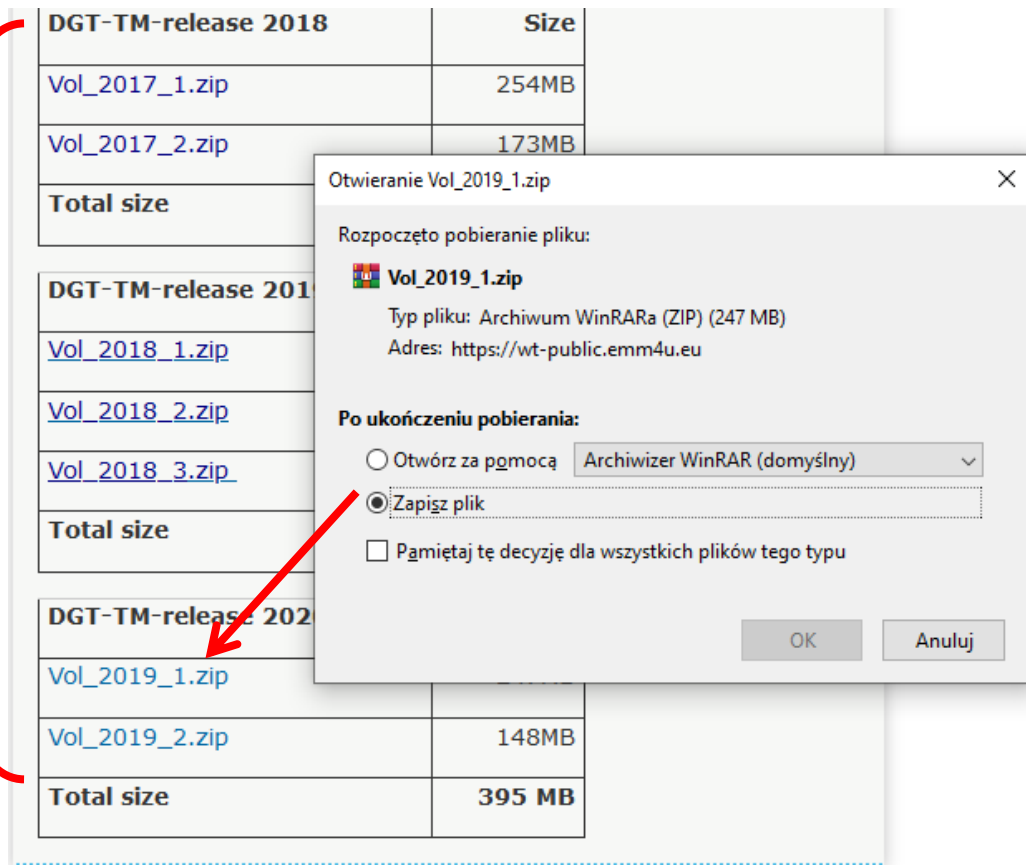
Vol_2019_1.zip	247MB
Vol_2019_2.zip	148MB
Total size	395 MB

DGT-TM-release 2021	Size
Vol_2020_1.zip	127MB
Vol_2020_2.zip	126MB
Vol_2020_3.zip	112MB
Vol_2020_4.zip	111MB
Vol_2020_5.zip	107MB
Total size	583MB

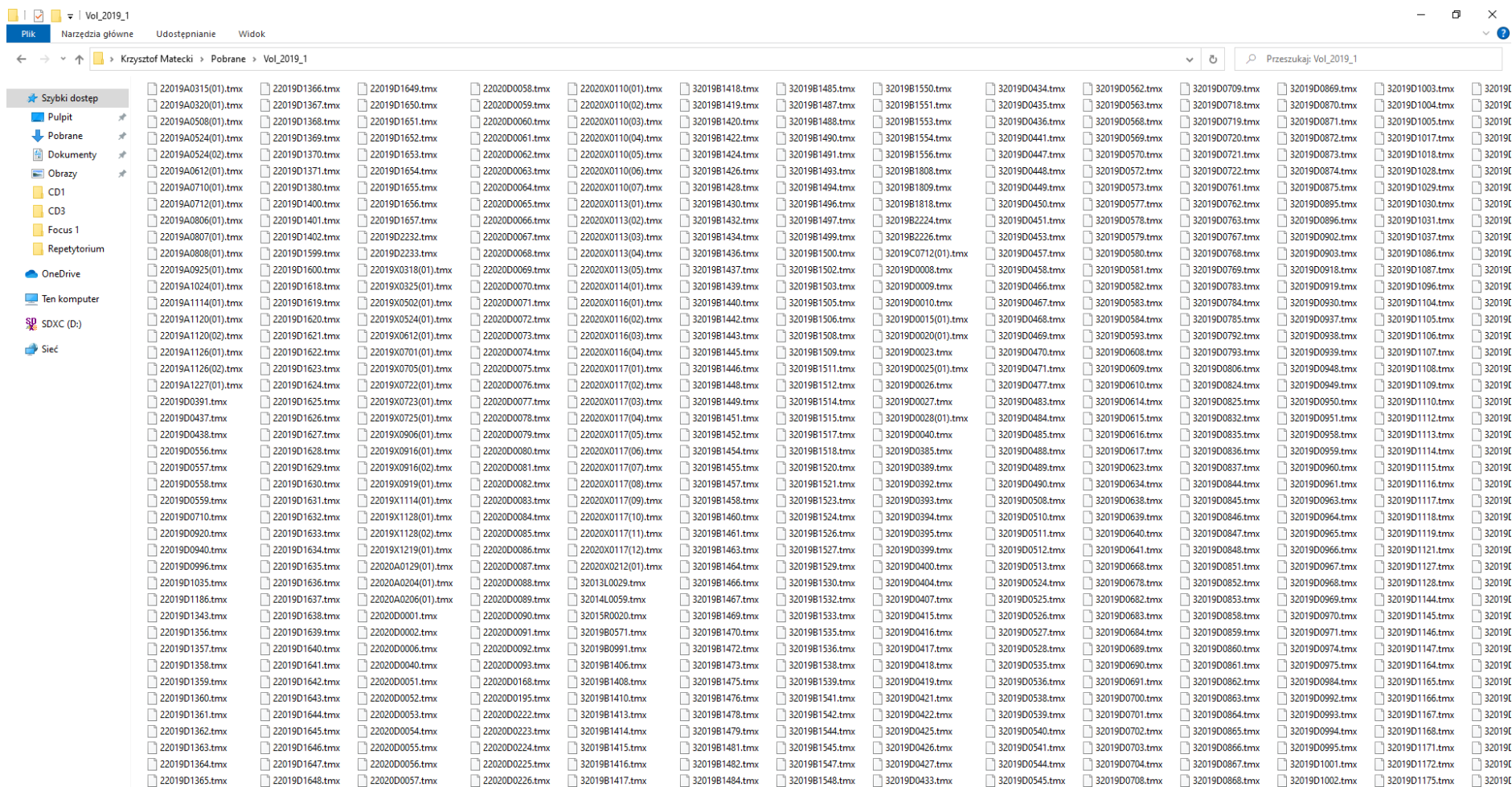
Pobieranie zarchiwizowanych danych pamięci DGT-TM

Pobrane pliki należy rozpakować, ażeby uniknąć problemów związanych z dostępem do ich zawartości przez program TMXtract.

W celu utworzenia kompletnego korpusu równoległego należy pobrać, rozpakować i scalić wszystkie archiwa.

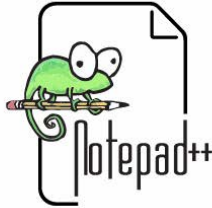


Zawartość archiwum Vol_2019_1



Podgląd zawartości jednego archiwum DGT-TM Vol_2019_1 – łącznie 1088 plików TMX

„Notepad ++” – edytor danych i kodów źródłowych



Current Version 8.1.4

- 🔖 [Home](#)
- 🔖 [Download](#)
- 🔖 [News](#)
- 🔖 [Online Help](#)
- 🔖 [Resources](#)
- 🔖 [RSS](#)
- 🔖 [Donate](#)
- 🔖 [Author](#)



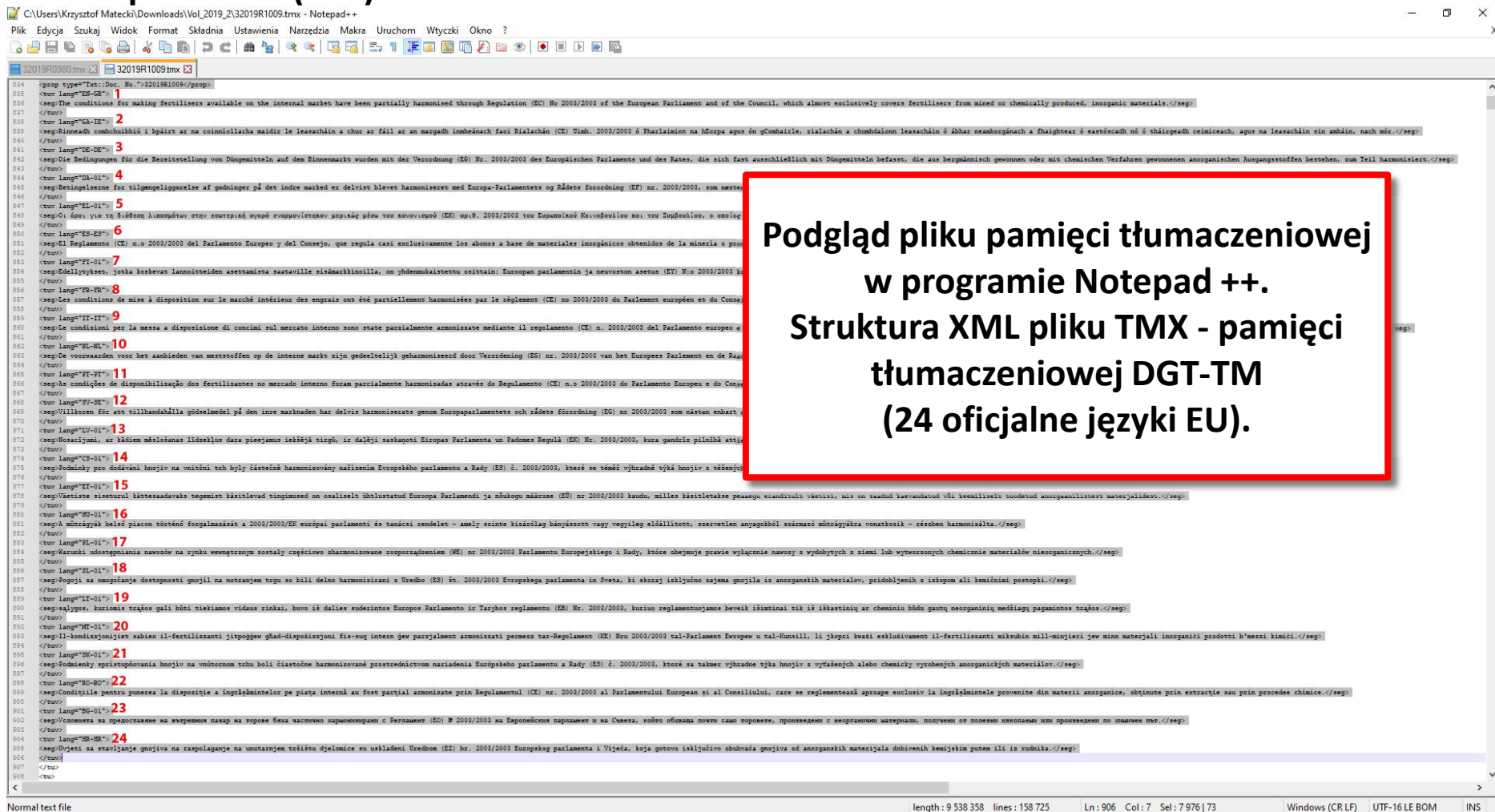
- [Installer](#) | [GPG Signature](#)
- [Portable \(zip\)](#) | [GPG Signature](#)
- [Portable \(7z\)](#) | [GPG Signature](#)
- [Mini-portable \(7z\)](#) | [GPG Signature](#)

Notepad++ jest rozbudowanym edytorem tekstu rozpowszechnianym na licencji GNU GPL (licencja wolnego i otwartego oprogramowania). Program posiada wiele przydatnych narzędzi pozwalających na wygodną edycję danych, m.in. edycję tabulatorów.

<https://notepad-plus-plus.org/downloads/>

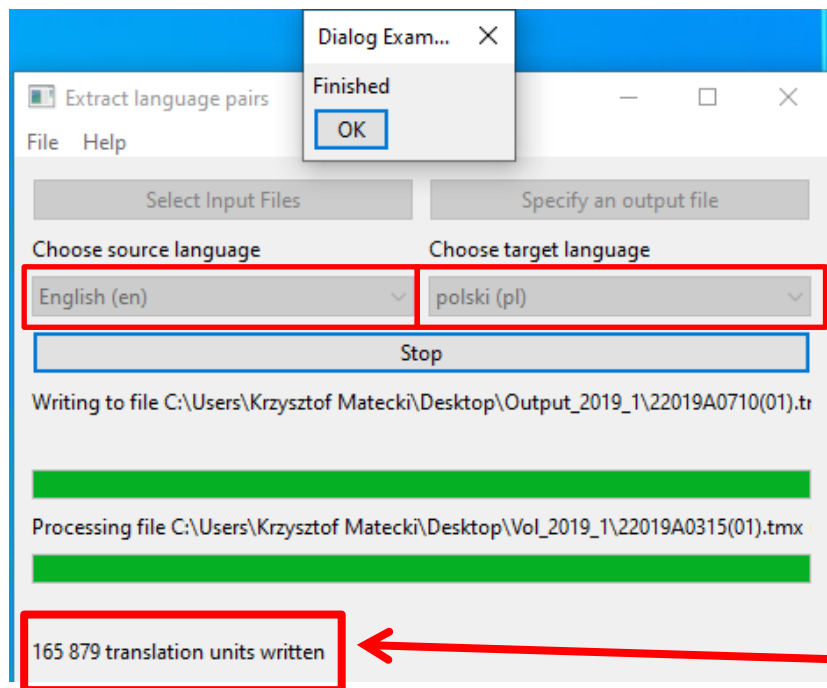
Zawartość plików - pamięci DGT-TM

Plik TMX (Translation Memory Exchange (TMX)) jest oparty na oparty na języku XML - standardzie stosowanym do wymiany pamięci tłumaczeń (TM). Dane TM są wykorzystywane przez pakiety oprogramowania do tłumaczenia wspomaganego komputerowo (CAT).



Program TMXtract – program scalający archiwa DGT-TM

Pamięć tłumaczeniowa DGT-TM jest zasobem językowym, któremu towarzyszy program TMXtract, umożliwiający tworzenie dwujęzycznych korpusów dla dowolnej pary językowej.



Program TMXtract wymaga zainstalowania na komputerze oprogramowania Java w wersji 32-bitowej.

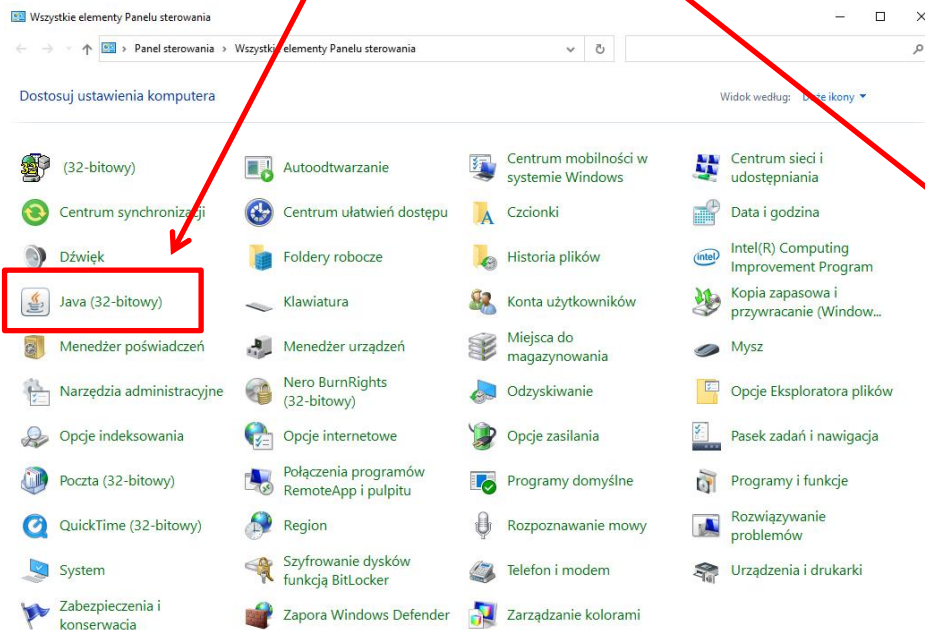


Program utworzył 165 879 bilingwalnych jednostek tekstu w jednym pliku.

Pobieranie 32-bitowej platformy Java

<https://www.java.com/en/download/manual.jsp>

Wersja 32-bitowa platformy Java dla 64-bitowych systemów operacyjnych Windows jest dostępna w spersonalizowanych ustawieniach pobierania.

A screenshot of the Java website's download page. The header is red with the Java logo and navigation links: 'Download', 'Help', 'Developers'. A search bar is in the top right. On the left, a 'Help Resources' sidebar lists links like 'Troubleshoot Java', 'Where can I get Java 7?', and 'Looking for the JDK?'. The main content area is titled 'Java Downloads for All Operating Systems' and 'Recommended Version 8 Update 301'. It features a yellow box with an 'Important Oracle Java License Update' notice. Below this, there's a section for selecting the download file based on the operating system. A red arrow points from the 'Java (32-bitowy)' icon in the Windows Control Panel to the 'Windows Offline (64-bit)' download option in the table.

Java Downloads for All Operating Systems
Recommended Version 8 Update 301
Release date July 20, 2021

Important Oracle Java License Update
The Oracle Java License has changed for releases starting April 16, 2019.
The new [Oracle Technology Network License Agreement for Oracle Java SE](#) is substantially different from prior Oracle Java licenses. The new license permits certain uses, such as personal use and development use, at no cost – but other uses authorized under prior Oracle Java licenses may no longer be available. Please review the terms carefully before downloading and using this product. An FAQ is available [here](#).
Commercial license and support is available with a low cost [Java SE Subscription](#).
Oracle also provides the latest OpenJDK release under the open source [GPL License](#) at [jdk.java.net](#).

Select the file according to your operating system from the list below to get the latest Java for your computer.

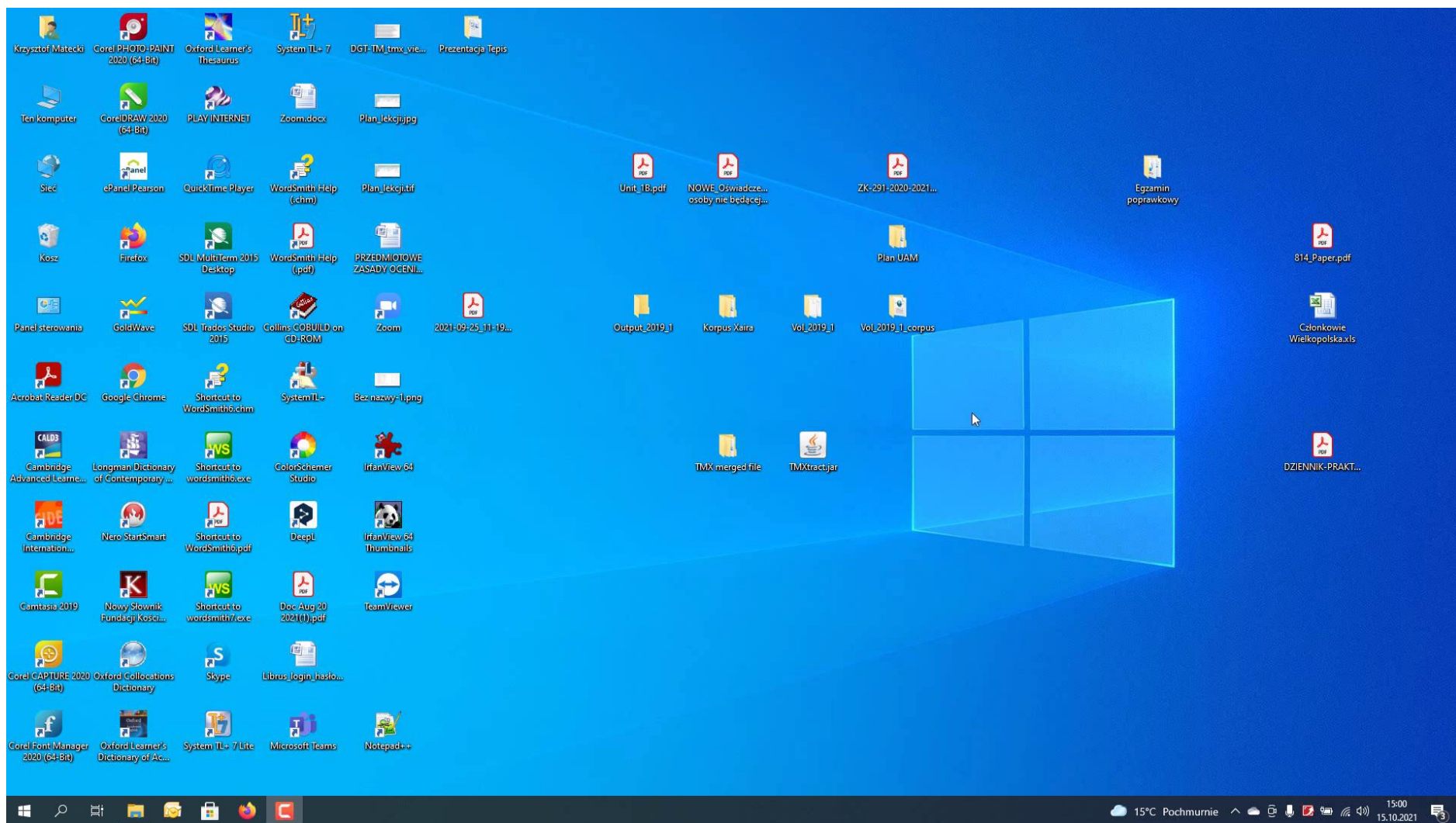
[All Java Downloads](#) [Remove Older Versions](#) [What is Java?](#)

By downloading Java you acknowledge that you have read and accepted the terms of the [Oracle Technology Network License Agreement for Oracle Java SE](#)

Windows		Which should I choose?	
Windows Online	filesize: 2 MB	Instructions	After installing Java, you may need to restart your browser in order to enable Java in your browser.
Windows Offline	filesize: 70.72 MB	Instructions	
Windows Offline (64-bit)	filesize: 81.08 MB	Instructions	

If you use 32-bit and 64-bit browsers interchangeably, you will need to install both 32-bit and 64-bit Java in order to have the Java plug-in for both browsers. » [FAQ about 64-bit Java for Windows](#)

Scalanie plików archiwum pamięci tłumaczeniowej DGT-TM - wideo



Struktura danych w pliku TMX

Jednoski tekstu EN-PL w scalonym pliku TMX – struktura XML

Jednostki te edytujemy do formatu tekstowego dla oprogramowania XAIRA.

Diagram illustrating the structure of a TMX file, showing the XML header and the body containing text segments.

Diagram Labels:

- identyfikator aktu prawnego (Legal Act Identifier)
- CELEX number: 3 2000 L 0060
- Sector
- Year
- Doc Type
- Doc Nr.

XML Header:

```
<?xml version="1.0" ?>
<!DOCTYPE tmx SYSTEM "tmx11.dtd">
<tmx version="version.1.1">
<header>
  <creationtool>TMXtract 1.2 26-09-2021</creationtool>
  <adminlang>EN-US</adminlang>
  <srclang>EN-GB</srclang>
</header>
<body>
```

Text Segments:

Segment 1 (EN-GB):

```
<prop type="Text::Doc. No.">22019A0320 (01)</prop>
<tuv xml:lang="EN-GB">
  <seg>Sustainable Fisheries Partnership Agreement between the European Union and the Kingdom of Morocco</seg>
</tuv>
<tuv xml:lang="PL-01">
  <seg>Umowa o partnerstwie w sprawie zrównoważonych połowów między Unią Europejską a Królestwem Marokańskim</seg>
</tuv>
```

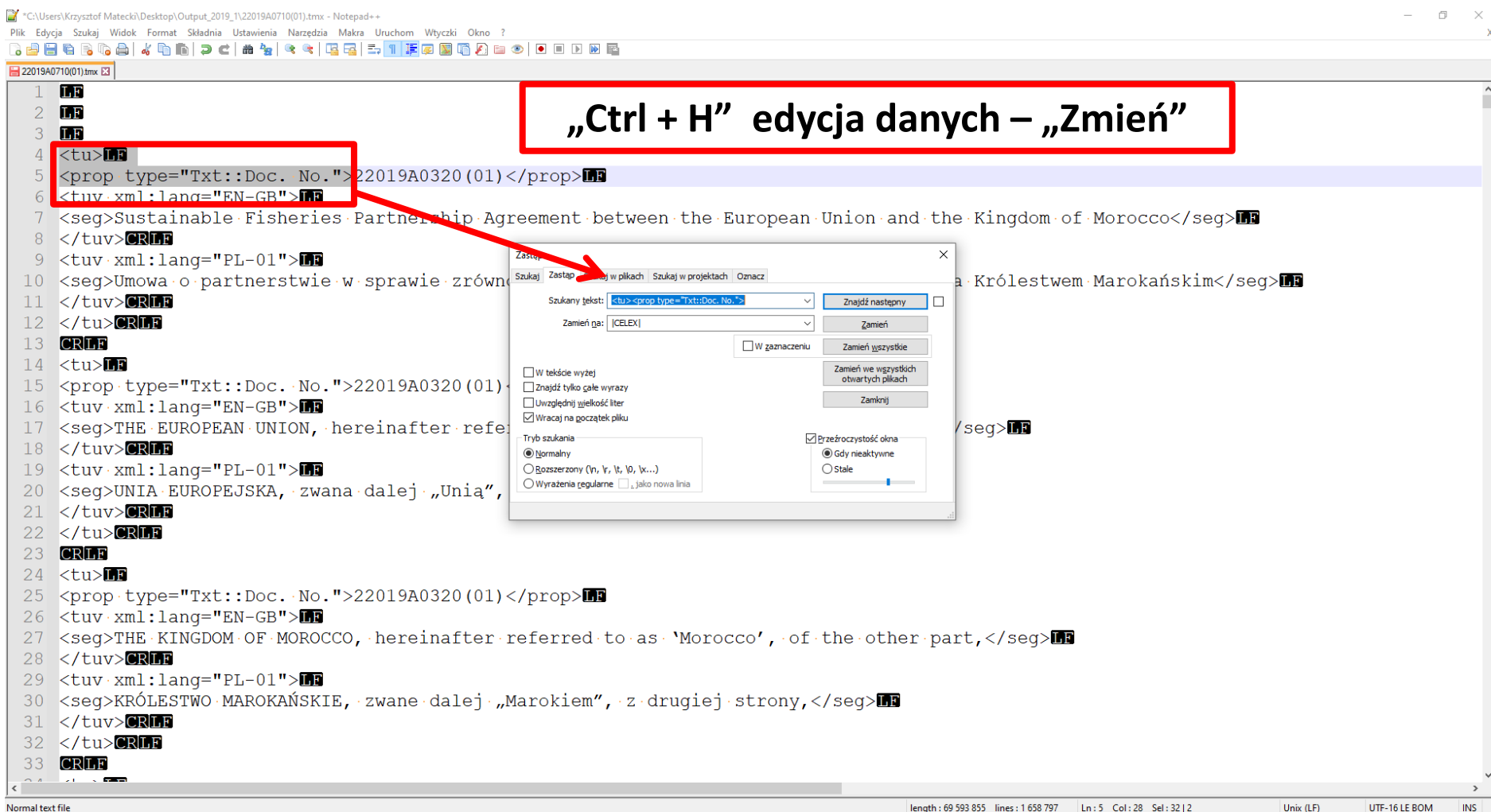
Segment 2 (EN-GB):

```
<prop type="Text::Doc. No.">22019A0320 (01)</prop>
<tuv xml:lang="EN-GB">
  <seg>THE EUROPEAN UNION, hereinafter referred to as 'the Union', of the one part,</seg>
</tuv>
<tuv xml:lang="PL-01">
  <seg>UNIA EUROPEJSKA, zwana dalej „Unią”, z jednej strony,</seg>
</tuv>
```

Normal text file

length : 69 594 044 lines : 1 658 805 Ln : 1 Col : 1 Pos : 1 Unix (LF) UTF-16 LE BOM INS

Edycja danych w pliku TMX



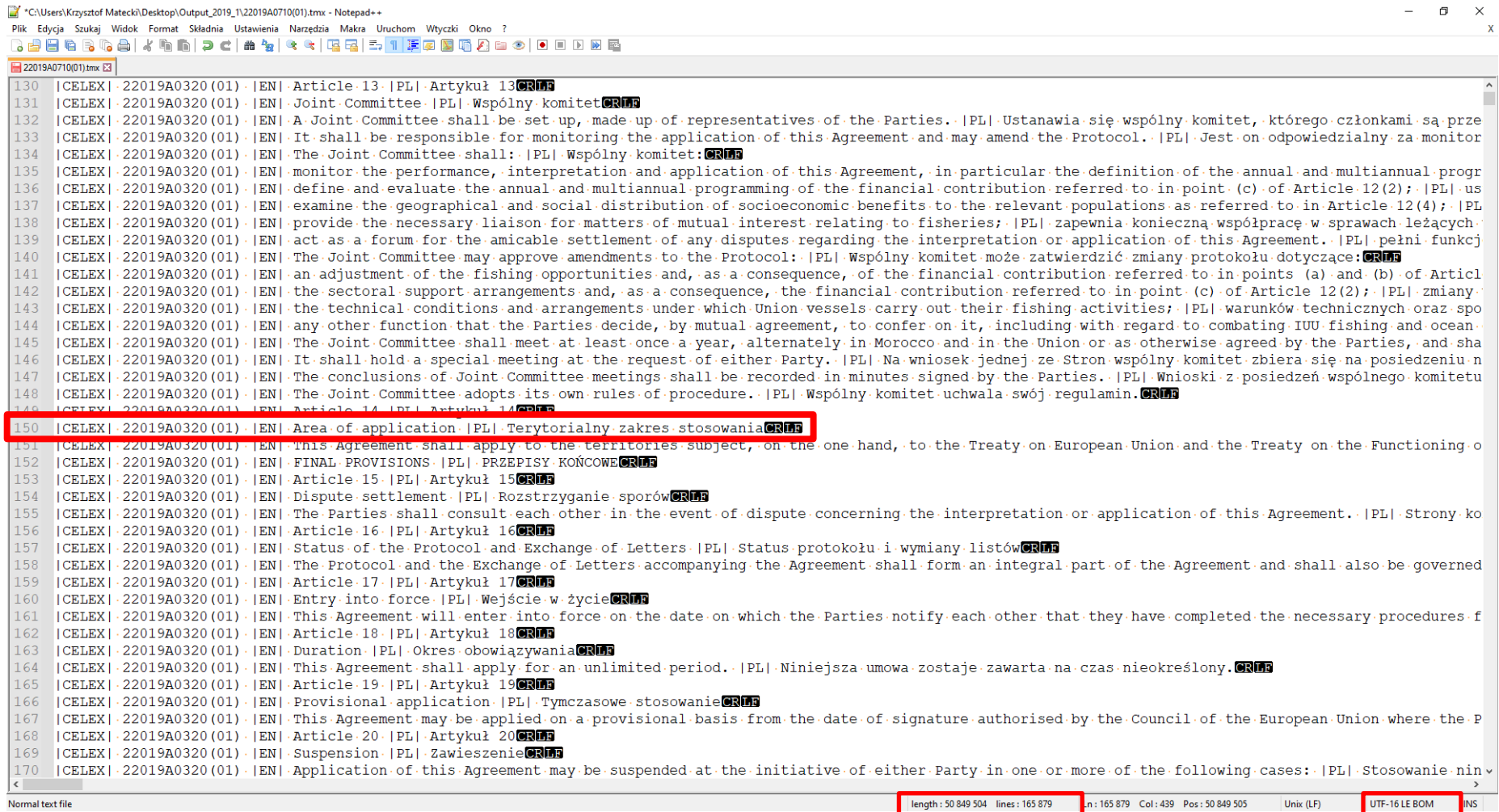
Edycja danych w pliku TMX - wideo

```
C:\Users\Krzysztof Matecki\Desktop\TMX merged file\TMX output\Vol_2019_1.tmx - Notepad++
Plik Edycja Szukaj Widok Format Składnia Ustawienia Narzędzia Makra Uruchom Wtyczki Okno ?
Vol_2019_1.tmx
1 <?xml version="1.0" ?>LF
2 <!DOCTYPE tmx SYSTEM "tmx11.dtd">LF
3 <tmx version="version.1.1">LF
4 <headerLF
5   <creationtool="TMXtract.1.2.15-10-2021"LF
6   <adminlang="EN-US"LF
7   <srclang="EN-GB"LF
8 >LF
9 </header>LF
10 LF
11 <body>LF
12 <tu>LF
13 <prop type="Txt::Doc.No.">22019A0320(01)</prop>LF
14 <tuv xml:lang="EN-GB">LF
15 <seg>Sustainable Fisheries Partnership Agreement between the European Union and the Kingdom of Morocco</seg>LF
16 </tuv>CRLF
17 <tuv xml:lang="PL-01">LF
18 <seg>Umowa o partnerstwie w sprawie zrównoważonych połowów między Unią Europejską a Królestwem Marokańskim</seg>LF
19 </tuv>CRLF
20 </tu>CRLF
21 CRLF
22 <tu>LF
23 <prop type="Txt::Doc.No.">22019A0320(01)</prop>LF
24 <tuv xml:lang="EN-GB">LF
25 <seg>THE EUROPEAN UNION, hereinafter referred to as 'the Union', of the one part,</seg>LF
26 </tuv>CRLF
27 <tuv xml:lang="PL-01">LF
28 <seg>UNIA EUROPEJSKA, zwana dalej „Unią”, z jednej strony,</seg>LF
29 </tuv>CRLF
30 </tu>CRLF
31 CRLF
32 <tu>LF
33 <prop type="Txt::Doc.No.">22019A0320(01)</prop>LF
34 <tuv xml:lang="EN-GB">LF
35 <seg>THE KINGDOM OF MOROCCO, hereinafter referred to as 'Morocco', of the other part,</seg>LF
36 </tuv>CRLF
37 <tuv xml:lang="PL-01">LF
38 <seg>KRÓLESTWO MAROKAŃSKIE, zwane dalej „Marokiem”, z drugiej strony,</seg>LF
39 </tuv>CRLF
40 </tu>CRLF
```

Normal text file length: 69 594 044 lines: 1 658 805 Ln: 1 Col: 1 Pos: 1 Unix (LF) UTF-16 LE BOM INS

15:10 15.10.2021 14°C Pochmurnie

Dane tekstowe po edycji w programie Notepad ++

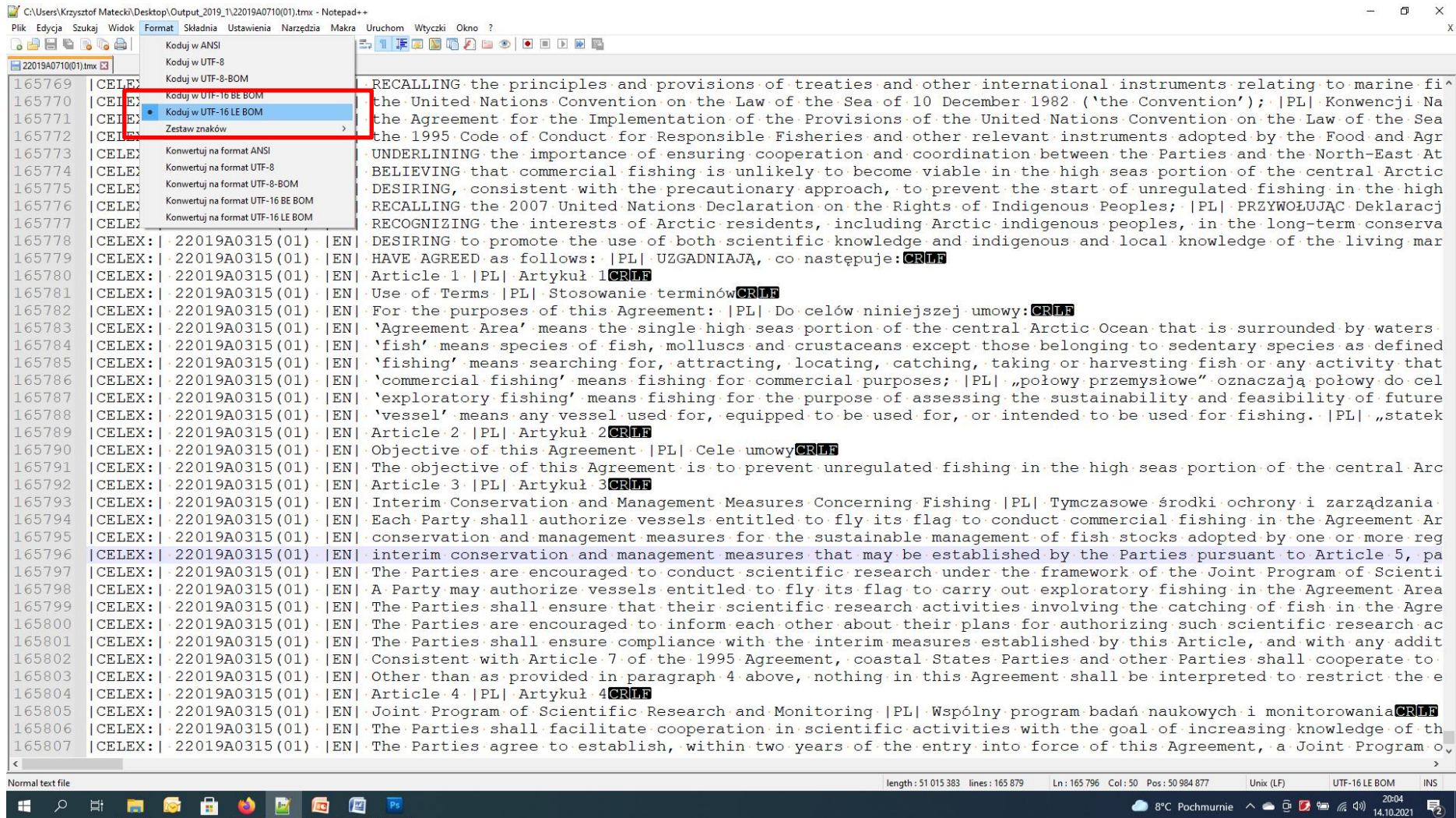


```
*C:\Users\Krzysztof Matecki\Desktop\Output_2019_1\22019A0710(01).tmx - Notepad++
Plik Edycja Szukaj Widok Format Składnia Ustawienia Narzędzia Makra Uruchom Wtyczki Okno ?
22019A0710(01).tmx
130 |CELEX| 22019A0320(01) |EN| Article 13. |PL| Artykuł 13. |CR|
131 |CELEX| 22019A0320(01) |EN| Joint Committee. |PL| Wspólny komitet. |CR|
132 |CELEX| 22019A0320(01) |EN| A Joint Committee shall be set up, made up of representatives of the Parties. |PL| Ustanawia się wspólny komitet, którego członkami są prze
133 |CELEX| 22019A0320(01) |EN| It shall be responsible for monitoring the application of this Agreement and may amend the Protocol. |PL| Jest on odpowiedzialny za monitor
134 |CELEX| 22019A0320(01) |EN| The Joint Committee shall: |PL| Wspólny komitet: |CR|
135 |CELEX| 22019A0320(01) |EN| monitor the performance, interpretation and application of this Agreement, in particular the definition of the annual and multiannual progr
136 |CELEX| 22019A0320(01) |EN| define and evaluate the annual and multiannual programming of the financial contribution referred to in point (c) of Article 12(2); |PL| us
137 |CELEX| 22019A0320(01) |EN| examine the geographical and social distribution of socioeconomic benefits to the relevant populations as referred to in Article 12(4); |PL|
138 |CELEX| 22019A0320(01) |EN| provide the necessary liaison for matters of mutual interest relating to fisheries; |PL| zapewnia konieczną współpracę w sprawach leżących
139 |CELEX| 22019A0320(01) |EN| act as a forum for the amicable settlement of any disputes regarding the interpretation or application of this Agreement. |PL| pełni funkcj
140 |CELEX| 22019A0320(01) |EN| The Joint Committee may approve amendments to the Protocol: |PL| Wspólny komitet może zatwierdzić zmiany protokołu dotyczące: |CR|
141 |CELEX| 22019A0320(01) |EN| an adjustment of the fishing opportunities and, as a consequence, of the financial contribution referred to in points (a) and (b) of Article
142 |CELEX| 22019A0320(01) |EN| the sectoral support arrangements and, as a consequence, the financial contribution referred to in point (c) of Article 12(2); |PL| zmiany
143 |CELEX| 22019A0320(01) |EN| the technical conditions and arrangements under which Union vessels carry out their fishing activities; |PL| warunków technicznych oraz spo
144 |CELEX| 22019A0320(01) |EN| any other function that the Parties decide, by mutual agreement, to confer on it, including with regard to combating IUU fishing and ocean
145 |CELEX| 22019A0320(01) |EN| The Joint Committee shall meet at least once a year, alternately in Morocco and in the Union or as otherwise agreed by the Parties, and sha
146 |CELEX| 22019A0320(01) |EN| It shall hold a special meeting at the request of either Party. |PL| Na wniosek jednej ze Stron wspólny komitet zbiera się na posiedzeniu n
147 |CELEX| 22019A0320(01) |EN| The conclusions of Joint Committee meetings shall be recorded in minutes signed by the Parties. |PL| Wnioski z posiedzeń wspólnego komitetu
148 |CELEX| 22019A0320(01) |EN| The Joint Committee adopts its own rules of procedure. |PL| Wspólny komitet uchwała swój regulamin. |CR|
149 |CELEX| 22019A0320(01) |EN| Article 14. |PL| Artykuł 14. |CR|
150 |CELEX| 22019A0320(01) |EN| Area of application. |PL| Terytorialny zakres stosowania |CR|
151 |CELEX| 22019A0320(01) |EN| This Agreement shall apply to the territories subject, on the one hand, to the Treaty on European Union and the Treaty on the Functioning o
152 |CELEX| 22019A0320(01) |EN| FINAL PROVISIONS. |PL| PRZEPISY KOŃCOWE |CR|
153 |CELEX| 22019A0320(01) |EN| Article 15. |PL| Artykuł 15. |CR|
154 |CELEX| 22019A0320(01) |EN| Dispute settlement. |PL| Rozstrzyganie sporów |CR|
155 |CELEX| 22019A0320(01) |EN| The Parties shall consult each other in the event of dispute concerning the interpretation or application of this Agreement. |PL| Strony ko
156 |CELEX| 22019A0320(01) |EN| Article 16. |PL| Artykuł 16. |CR|
157 |CELEX| 22019A0320(01) |EN| Status of the Protocol and Exchange of Letters. |PL| Status protokołu i wymiany listów |CR|
158 |CELEX| 22019A0320(01) |EN| The Protocol and the Exchange of Letters accompanying the Agreement shall form an integral part of the Agreement and shall also be governed
159 |CELEX| 22019A0320(01) |EN| Article 17. |PL| Artykuł 17. |CR|
160 |CELEX| 22019A0320(01) |EN| Entry into force. |PL| Wejście w życie |CR|
161 |CELEX| 22019A0320(01) |EN| This Agreement will enter into force on the date on which the Parties notify each other that they have completed the necessary procedures f
162 |CELEX| 22019A0320(01) |EN| Article 18. |PL| Artykuł 18. |CR|
163 |CELEX| 22019A0320(01) |EN| Duration. |PL| Okres obowiązywania |CR|
164 |CELEX| 22019A0320(01) |EN| This Agreement shall apply for an unlimited period. |PL| Niniejsza umowa zostaje zawarta na czas nieokreślony. |CR|
165 |CELEX| 22019A0320(01) |EN| Article 19. |PL| Artykuł 19. |CR|
166 |CELEX| 22019A0320(01) |EN| Provisional application. |PL| Tymczasowe stosowanie |CR|
167 |CELEX| 22019A0320(01) |EN| This Agreement may be applied on a provisional basis from the date of signature authorised by the Council of the European Union where the P
168 |CELEX| 22019A0320(01) |EN| Article 20. |PL| Artykuł 20. |CR|
169 |CELEX| 22019A0320(01) |EN| Suspension. |PL| Zawieszenie |CR|
170 |CELEX| 22019A0320(01) |EN| Application of this Agreement may be suspended at the initiative of either Party in one or more of the following cases: |PL| Stosowanie nin
Normal text file
length: 50 849 504 lines: 165 879 in: 165 879 Col: 439 Pos: 50 849 505 Unix (LF) UTF-16 LE BOM INS
```

Liczba słów: 50 849 504; liczba wierszy 165 879; kodowanie UTF-16 LE BOM

Kodowanie czcionek => UNICODE

Standard UNICODE to system kodowania znaków, który pozwala na stosowanie symboli, w tym także znaków diakrytycznych, istniejących we wszystkich językach świata. UTF-16 (ang. *16-bit Unicode Transformation Format*) jest jednym ze sposobów kodowania znaków w standardzie Unicode.



The screenshot shows a Notepad++ window with a text file open. The menu is open, and the 'Koduj w UTF-16 LE BOM' option is selected. The text in the background is a legal document in English and Polish, discussing the United Nations Convention on the Law of the Sea and fishing regulations. The status bar at the bottom indicates the file is in 'Normal text file' format, with a length of 51 015 383 bytes, 165 879 lines, and is currently in 'UTF-16 LE BOM' encoding.

```
165769 | CELEX: | 22019A0315 (01) | EN | RECALLING the principles and provisions of treaties and other international instruments relating to marine fi
165770 | CELEX: | 22019A0315 (01) | EN | the United Nations Convention on the Law of the Sea of 10 December 1982 ('the Convention'); | PL | Konwencji Na
165771 | CELEX: | 22019A0315 (01) | EN | the Agreement for the Implementation of the Provisions of the United Nations Convention on the Law of the Sea
165772 | CELEX: | 22019A0315 (01) | EN | the 1995 Code of Conduct for Responsible Fisheries and other relevant instruments adopted by the Food and Agr
165773 | CELEX: | 22019A0315 (01) | EN | UNDERLINING the importance of ensuring cooperation and coordination between the Parties and the North-East At
165774 | CELEX: | 22019A0315 (01) | EN | BELIEVING that commercial fishing is unlikely to become viable in the high seas portion of the central Arctic
165775 | CELEX: | 22019A0315 (01) | EN | DESIRING, consistent with the precautionary approach, to prevent the start of unregulated fishing in the high
165776 | CELEX: | 22019A0315 (01) | EN | RECALLING the 2007 United Nations Declaration on the Rights of Indigenous Peoples; | PL | PRZYZWOŁUJĄC Deklaracj
165777 | CELEX: | 22019A0315 (01) | EN | RECOGNIZING the interests of Arctic residents, including Arctic indigenous peoples, in the long-term conserva
165778 | CELEX: | 22019A0315 (01) | EN | DESIRING to promote the use of both scientific knowledge and indigenous and local knowledge of the living mar
165779 | CELEX: | 22019A0315 (01) | EN | HAVE AGREED as follows: | PL | UZGADNIAJĄ, co następuje: CRIF
165780 | CELEX: | 22019A0315 (01) | EN | Article 1 | PL | Artykuł 1 CRIF
165781 | CELEX: | 22019A0315 (01) | EN | Use of Terms | PL | Stosowanie terminów CRIF
165782 | CELEX: | 22019A0315 (01) | EN | For the purposes of this Agreement: | PL | Do celów niniejszej umowy: CRIF
165783 | CELEX: | 22019A0315 (01) | EN | 'Agreement Area' means the single high seas portion of the central Arctic Ocean that is surrounded by waters
165784 | CELEX: | 22019A0315 (01) | EN | 'fish' means species of fish, molluscs and crustaceans except those belonging to sedentary species as defined
165785 | CELEX: | 22019A0315 (01) | EN | 'fishing' means searching for, attracting, locating, catching, taking or harvesting fish or any activity that
165786 | CELEX: | 22019A0315 (01) | EN | 'commercial fishing' means fishing for commercial purposes; | PL | „połowy przemysłowe” oznaczają połowy do cel
165787 | CELEX: | 22019A0315 (01) | EN | 'exploratory fishing' means fishing for the purpose of assessing the sustainability and feasibility of future
165788 | CELEX: | 22019A0315 (01) | EN | 'vessel' means any vessel used for, equipped to be used for, or intended to be used for fishing. | PL | „statek
165789 | CELEX: | 22019A0315 (01) | EN | Article 2 | PL | Artykuł 2 CRIF
165790 | CELEX: | 22019A0315 (01) | EN | Objective of this Agreement | PL | Cele umowy CRIF
165791 | CELEX: | 22019A0315 (01) | EN | The objective of this Agreement is to prevent unregulated fishing in the high seas portion of the central Arc
165792 | CELEX: | 22019A0315 (01) | EN | Article 3 | PL | Artykuł 3 CRIF
165793 | CELEX: | 22019A0315 (01) | EN | Interim Conservation and Management Measures Concerning Fishing | PL | Tymczasowe środki ochrony i zarządzania
165794 | CELEX: | 22019A0315 (01) | EN | Each Party shall authorize vessels entitled to fly its flag to conduct commercial fishing in the Agreement Ar
165795 | CELEX: | 22019A0315 (01) | EN | conservation and management measures for the sustainable management of fish stocks adopted by one or more reg
165796 | CELEX: | 22019A0315 (01) | EN | interim conservation and management measures that may be established by the Parties pursuant to Article 5, pa
165797 | CELEX: | 22019A0315 (01) | EN | The Parties are encouraged to conduct scientific research under the framework of the Joint Program of Scienti
165798 | CELEX: | 22019A0315 (01) | EN | A Party may authorize vessels entitled to fly its flag to carry out exploratory fishing in the Agreement Area
165799 | CELEX: | 22019A0315 (01) | EN | The Parties shall ensure that their scientific research activities involving the catching of fish in the Agre
165800 | CELEX: | 22019A0315 (01) | EN | The Parties are encouraged to inform each other about their plans for authorizing such scientific research ac
165801 | CELEX: | 22019A0315 (01) | EN | The Parties shall ensure compliance with the interim measures established by this Article, and with any addit
165802 | CELEX: | 22019A0315 (01) | EN | Consistent with Article 7 of the 1995 Agreement, coastal States Parties and other Parties shall cooperate to
165803 | CELEX: | 22019A0315 (01) | EN | Other than as provided in paragraph 4 above, nothing in this Agreement shall be interpreted to restrict the e
165804 | CELEX: | 22019A0315 (01) | EN | Article 4 | PL | Artykuł 4 CRIF
165805 | CELEX: | 22019A0315 (01) | EN | Joint Program of Scientific Research and Monitoring | PL | Wspólny program badań naukowych i monitorowania CRIF
165806 | CELEX: | 22019A0315 (01) | EN | The Parties shall facilitate cooperation in scientific activities with the goal of increasing knowledge of th
165807 | CELEX: | 22019A0315 (01) | EN | The Parties agree to establish, within two years of the entry into force of this Agreement, a Joint Program o
```

Oprogramowanie korpusowe XAIRA

<https://xaira.soft112.com/>

Welcome to XAIRA on sourceforge

Users

[Download Xaira files and releases](#)

[Donate money](#)

[Read project details](#)

[Get support](#)

[Read background documentation and information about the project](#)



What is XAIRA?

About this project:

This project is responsible for development and maintenance of XAIRA (XML Aware Indexing and Retrieval Architecture) an open source software package which supports indexing and analysis of large XML textual resources such as natural language corpora.

The project was registered on SourceForge.net on Feb 1, 2005.

About SourceForge.net:

[SourceForge.net](#) is the world's largest provider of hosting for Open Source software development projects. SourceForge.net provides a variety of services to projects, including a download mirror network, collaborative development tools (like CVS and Subversion), and tools to support discussion and support. These services are provided to projects and their end-users free-of-charge.

About Open Source:

Of benefit to users, [Open Source](#) software is licensed so you can download and use the software free-of-charge. The source code for this software is made available free-of-charge, you (or a programmer you hire) can make changes to this software to better meet your needs, and you can release your changed code back to the community passing the benefit on to other users.

The license terms for XAIRA are specified in the licensing documents included with each [packaged release](#).

Oprogramowanie XAIRA

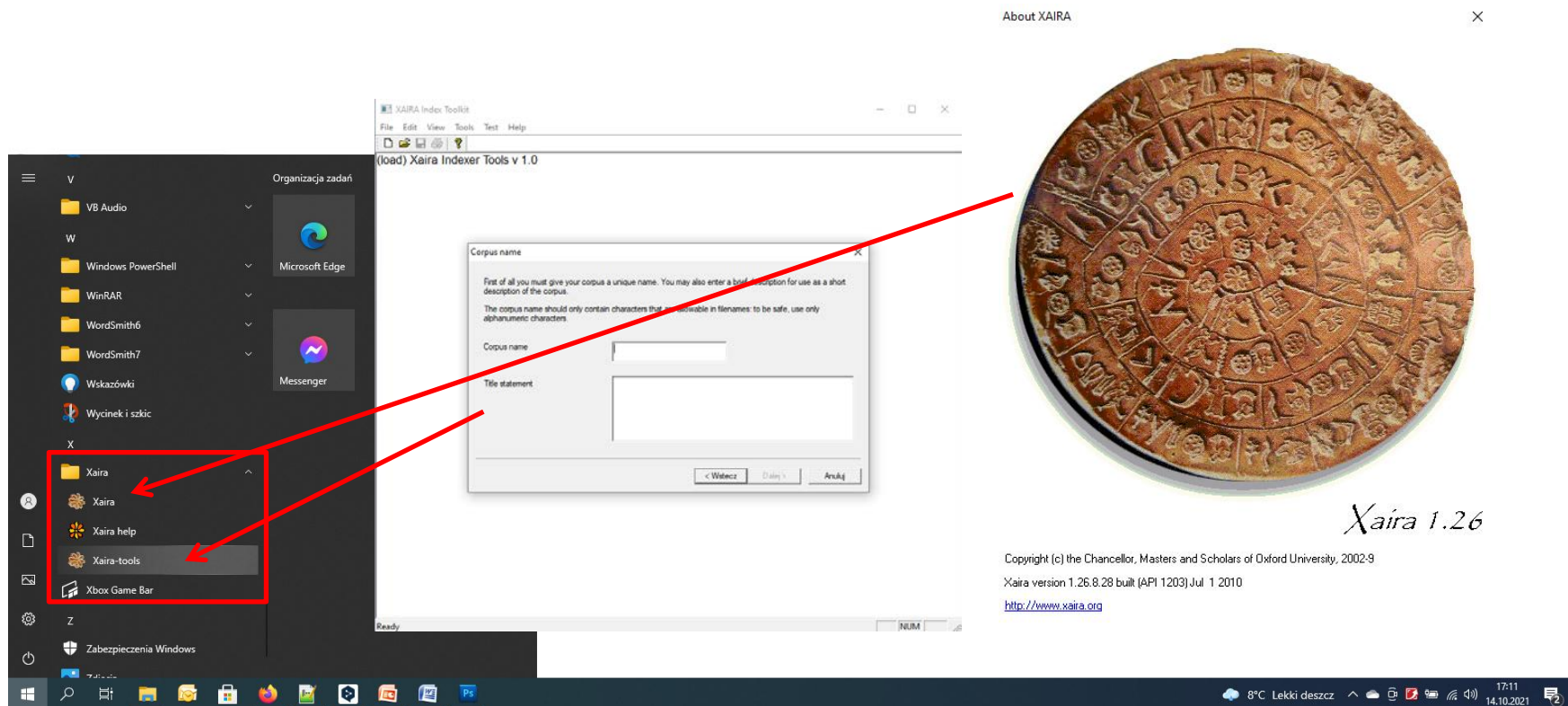
XAIRA (XML Aware Indexing and Retrieval Architecture), platforma stworzona w latach 2005-2006 na Uniwersytecie Oksfordzkim, jest oprogramowaniem *open source*, które umożliwia indeksowanie i analizę dużych danych tekstowych.

Przy pomocy oprogramowania można tworzyć z plików tekstowych pliki XML, które po zindeksowaniu umożliwiają szybkie wyszukiwanie danych w korpusie. Xaira jest oparta na systemie SARA, stworzonym do przeszukiwania Brytyjskiego Korpusu Narodowego.

XML (ang. Extensible Markup Language, rozszerzalny język znaczników) – uniwersalny język znaczników przeznaczony do reprezentowania różnych danych w strukturalizowany sposób. (ź. Wikipedia).

Architektura XAIRA

Oprogramowanie XAIRA składa się z dwóch programów: kompilatora (narzędzia indeksującego, *ang. Indexer Toolkit*) oraz wyszukiwarki korpusowej XAIRA.



The image displays a Windows 10 desktop environment. On the left, the Start menu is open, showing a list of applications. A red box highlights the 'Xaira' folder, which contains 'Xaira', 'Xaira help', and 'Xaira-tools'. A red arrow points from this folder to the 'XAIRA Indexer Toolkit' window. The 'XAIRA Indexer Toolkit' window is open, showing a 'Corpus name' dialog box. The dialog box contains instructions: 'First of all you must give your corpus a unique name. You may also enter a brief description for use as a short description of the corpus. The corpus name should only contain characters that are permissible in filenames: to be safe, use only alphanumeric characters.' It has input fields for 'Corpus name' and 'Title statement', and buttons for '< Wstecz', 'Dalej >', and 'Anuluj'. To the right of the dialog box is a large, circular, ancient-looking coin or medallion with intricate carvings. Below the coin, the text 'Xaira 1.26' is written in a stylized font. At the bottom right, there is copyright information: 'Copyright (c) the Chancellor, Masters and Scholars of Oxford University, 2002-9', 'Xaira version 1.26: 8.28 built (API 1203) Jul 1 2010', and the website 'http://www.xaira.org'. The system tray at the bottom shows the date and time as '17:11 14.10.2021'.

About XAIRA

XAIRA Indexer Toolkit

(load) Xaira Indexer Tools v 1.0

Corpus name

First of all you must give your corpus a unique name. You may also enter a brief description for use as a short description of the corpus.

The corpus name should only contain characters that are permissible in filenames: to be safe, use only alphanumeric characters

Corpus name

Title statement

< Wstecz Dalem > Anuluj

Xaira 1.26

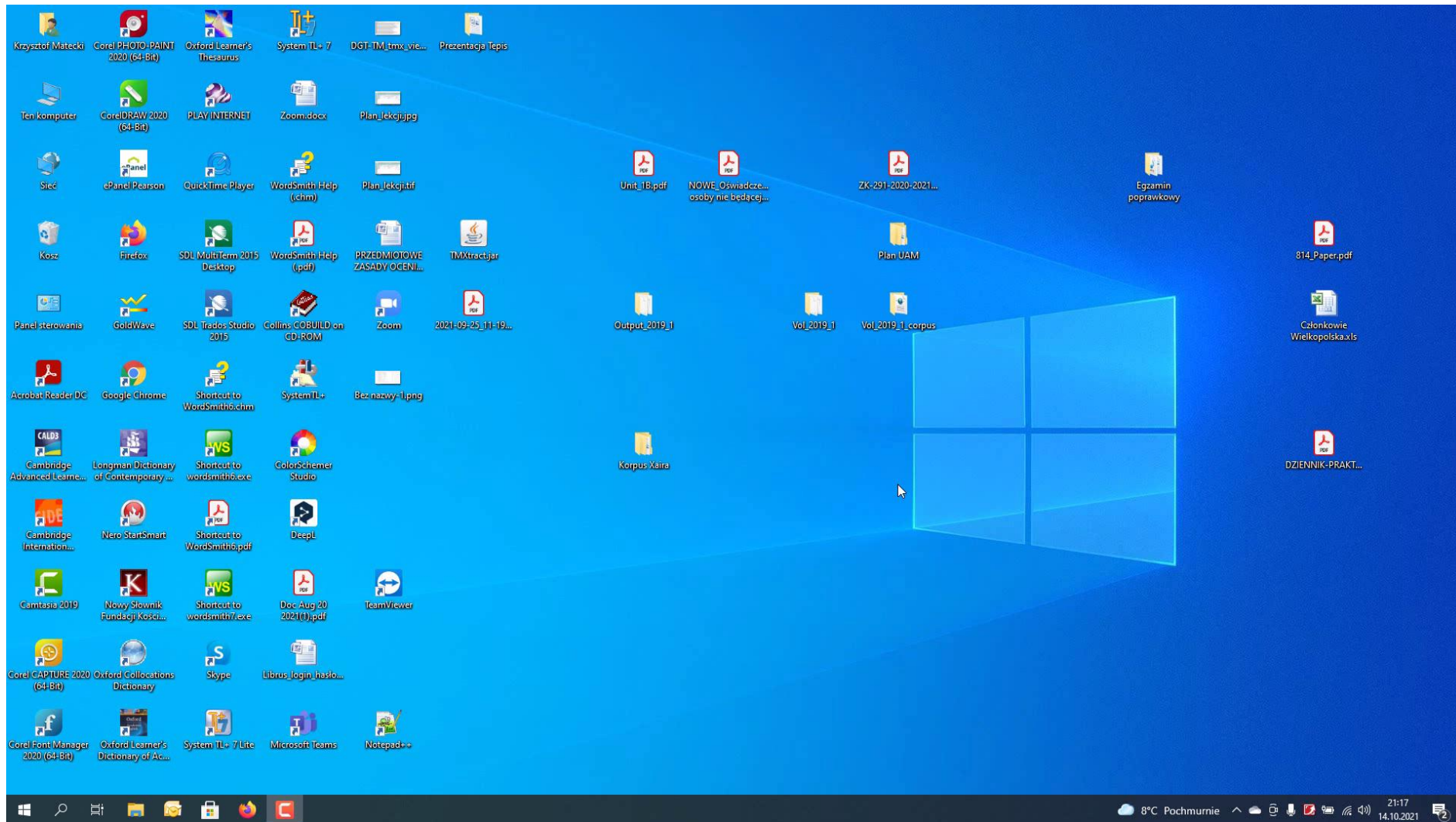
Copyright (c) the Chancellor, Masters and Scholars of Oxford University, 2002-9

Xaira version 1.26: 8.28 built (API 1203) Jul 1 2010

<http://www.xaira.org>

8°C Lekki deszcz 17:11 14.10.2021

Tworzenie (kompilacja) korpusu przy pomocy oprogramowania XAIRA - wideo



Kilka pytań do dyskusji i refleksji

1. Jak pracować wspólnie w czasach, gdy komputery i sztuczna inteligencja (AI) przejmują tak wiele zadań?
2. Czy automatyzacja tylko zwiększy poziom współpracy z maszynami, czy też specjaliści znajdą nowe sposoby na interakcję 'offline'? Jak te sposoby pracy wpłyną na pracę tłumacza?
3. Czy możemy oczekiwać, że będziemy pracować mniej i wydajniej? Czy też będziemy pracować jeszcze dłużej, edytując ogromne ilości tłumaczenia maszynowego?
4. Czy już żyjemy w epoce "przemysłu post-edytorskiego"?