Document Analysis (Extraction) Tool Instructions - copy 1

To make a login, click on "Register here"

Login

- You have been logged out.

Username:

Password:

Login

Don't have an account? Register here.

Enter your preferred username and password (both are case-sensitive). You will also need to enter a security Token in the third box. (Each security Token can be used to create a limited number of accounts, then it expires.)

Register

Username:

test_user

Password:

••••••••

Token:

43b2c764-62fb-47ce-af2b-5

Register

After registering you can login using your new username and password (you won't need a security Token anymore unless you want to create a new account).

After logging in, you will be able to upload your PDFs. Select "Choose Files" and select all the PDFs you want to upload, you can upload them all at once.

After closing the dialog box, the number of PDFs you want to upload will be displayed next to the button. You can then press the blue button, "Upload PDFs" to upload and process them.

The progress of each PDF upload will be displayed with a green progress bar:
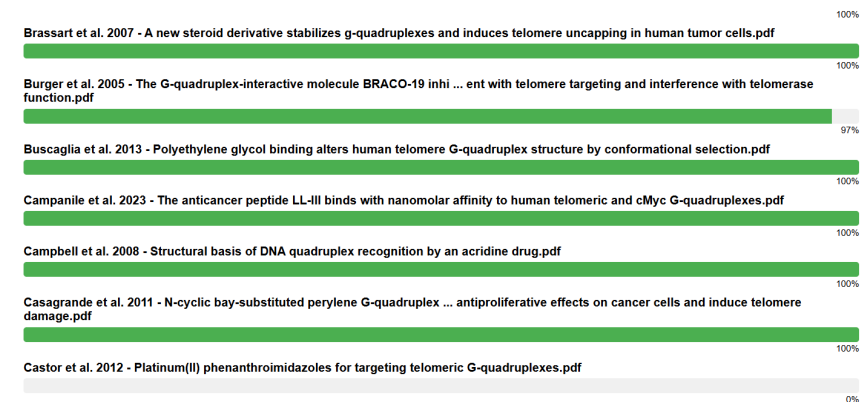
The text from the PDFs will then be automatically extracted, and PubMed ID (PMID) hopefully discovered along with bibliographic data. You can see the text extraction process below the progress bars:

After this process is complete, you will see a large table containing your PDFs and some of their bibliographic data. You can delete any PDF from this list (by clicking the "Delete"

button in the last column), or add new PDFs to your collection via the same uploading process as above. Your PDFs are stored in the server after you logout, so you don't need to upload them again.

Your PDFs

Filename	PMID	Title	Authors	Publication Year	Actions
Abdel-Rahman_and_Morsy_2007_-_Novel_benzothiazolyl_urea_and_thiourea_derivatives_with_potential_cytotoxic_and_antimicrobial_activities.pdf	17373548	Novel benzothiazolyl urea and thiourea derivatives with potential cytotoxic and antimicrobial activities	Hamdy M. Abdel-Rahman, Mohamed A. Morsy	2007	Delete
Alessandrini_et_al._2022_-_Telomere_as_a_therapeutic_target_in_dedifferentiated_liposarcoma.pdf	35681604	Telomere as a Therapeutic Target in Dedifferentiated Liposarcoma	Irene Alessandrini, Stefano Percio, Eisa Naghshineh, Valentina Zuco, Silvia Stacchiotti, Alessandro Gronchi, Sandro Pasquali, Nadia Zaffaroni, Marco Folini		Delete
Abou_Assi_et_al._2017_-_2-Fluoroarabinonucleic_acid_modification_traps_G-quadruplex_and_i-motif_structures_in_human_telomeric_DNA.pdf	29036537	2'-Fluoroarabinonucleic acid modification traps G-quadruplex and i-motif structures in human telomeric DNA	Hala Abou Assi, Roberto El-Khoury, Carlos González, Masad J. Damha	2017	Delete
Ali_et_al._2021_-_Pt-ttpy_a_G-quadruplex_binding_platinum_complex_induces_telomere_dysfunction_and_G-rich_regions_DNA_damage.pdf	34021581	Pt-ttpy, a G-quadruplex binding platinum complex, induces telomere dysfunction and G-rich regions DNA damage	Samar Ali, Emilia Puig Lombardi, Deepanjan Ghosh, Tao Jia, Géraldine Vitry, Lina Saker, Joël Poupon, Marie-Paule Teulade-Fichou, Alain Nicolas, Arturo Londono-Vallejo, Sophie Bombard	2021	Delete
Agrawal_et_al._2008_-_Energetics_of_the_human_Tel-22_quadruplex-telomestatin_interaction_-_a_molecular_dynamics_study.pdf	18461983	Energetics of the Human Tel-22 Quadruplex–Telomestatin Interaction: A Molecular Dynamics Study	Saurabh Agrawal, Rajendra Prasad Ojha, Souvik Maiti	2008	Delete

(Note that there is currently an arbitrary 1 million character limit on the length of the PDF. If the PDF exceeds this limit it won't show up in the table after uploading. To bypass this limit, you can break the PDF up into two or more separate PDFs, or isolate the pages of interest, and re-upload them.)

To ask specific extraction questions from your PDFs, scroll down to the box below this table, and enter each extraction question on its own separate line. You don't need to number them, the server will do it automatically.

Enter Prompts

```
Write a Shakespearean sonnet about the study's introduction and premise.
Write a haiku about the study's methods.
Write a Confucian proverb about the study's results.
Write a modernist poem about the study's discussion and conclusion.
```

Process PDFs with Prompts

You may also (optionally) provide "Additional Context," or information to pass to the AI to help it interpret the prompts and the manuscripts. To do so, you can simply cut and paste such information from your desired source(s) into this box under the "Notes" section. This could include definitions of terms or detailed descriptions of theories:

Additional Context (Optional):

```
You are an assistant tasked with responding to questions that require extracting
information from a **Manuscript Text** to be provided by the user.

# Steps
1. **Data Identification**: Review the user-provided **Manuscript Text**.
2. **Extraction Planning**: Review the user-provided questions and associated
instructions to reduce ambiguities and create a clear plan to apply them to the
**Manuscript Text**.
3. **Question Application**: Use your plan to apply each question to the
**Manuscript Text** to guide the identification of factual information.
3. **Fact Extraction**: Identify and document factual information based on the
questions. Prioritize verifiable facts over subjective opinions or
interpretations.
4. **Summarization**: Present the extracted facts in the format requested by the
user.

# Notes
- Ensure that all extracted information is verifiable and factual.
- Avoid including subjective statements or opinions from the qualitative data.
- **********ADD ANY ADDITONAL NOTES HERE ***********

# Manuscript Text Details
- The **Manuscript Text** you will be provided represents scholarly papers from
the field of health professions education, from journals such as Academic
Medicine, Medical Education, Advances in Health Sciences Education, Teaching and
Learning in Medicine, etc.
```

When you are ready to process these questions, please click, "Process PDFs with Prompts."

Now the server will use the AI to read each PDF and answer your questions from it. As each PDF is processed, the answers to your questions will populate automatically in the table below your questions:

Results

Download Results as Excel

Filename	PMID	Q1	Q2	Q3	Q4
Abdel-Rahman_an...	17373548	A sonnet highlighting the study's aim to combat cancer and microbial infections through novel benzothiazole urea and thiourea derivatives.	A haiku summarizing the synthesis and evaluation methods of the compounds.	A Confucian proverb reflecting on the comparative effectiveness of synthesized compounds.	A modernist poem reflecting on the study's outcomes and future research directions.
Alessandrini_et...	35681604	A sonnet about DDLPS's severity and the need for new treatments.	A haiku about comparing gene expression profiles in DDLPS.	A proverb about the wisdom of targeting telomeres in cancer treatment.	A modernist poem about RHPS4's effects on DDLPS cells and the promise of telomere-targeting therapies.
Abou_Assi_et_al	29036537	A sonnet about human telomeres' role in aging and cancer through the study of G-quadruplex	A haiku about synthesizing and purifying modified DNA	A proverb about the balance in controlling DNA structures through	A modernist poem about the complex interplay between DNA structures and

Note that processing can take some time (on average, just under one minute), especially if you have long PDFs and many questions. The website will inform you when it's done (i.e. see dialog below). **If you log out of the website or close the browser window during processing, your PDF extraction will stop**, so please keep the window open during processing. After processing, your extraction results will be saved on the website, so that

you can log in at any time to download them again. However, if you choose to process the prompts again, or process new prompts, your prior responses will be overwritten.
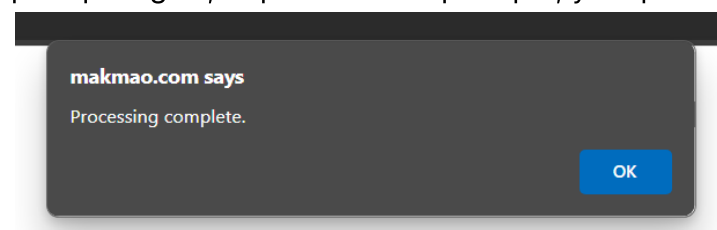
Note that the table only contains Short Answers to your questions. To receive the full answer, you can click on any cell of the table to read it and see the supporting quotes, Replica responses, as well as the certainty estimate and rationale:

You can download an Excel file with both the Long/Short Answers and any relevant quotes from each PDF. To download this Excel file, just click the "Download Results as Excel" at the top of this table. (You can download the Excel file anytime during processing, it will contain only the data processed so far.)

Our example's Excel result for the first prompt looks like this:

IF_FileNam	PMID	Q_1: Write a Shakespearean sonnet about the study's introduction and premise. - Explanation	Q_1: Write a Shakespearean sonnet about the study's introduction and premise. - Quote	Q_1: Write a Shakespearean sonnet about the study's introduction and premise. - Brief Answer
Ali_et_al._	34021581	In a realm where metals and molecules dance, A tale unfolds of Pt-ttpy's advance; Where G-quadruplexes form their stance, And telomeres face the platinum chance. From stable structures to dynamic play, These quadruplexes in cells sway. With potential sequences by the day, They're targets for complexes, come what may. The study seeks to find where Pt-ttpy's might Lies within human genome's intricate light; Telomere dysfunction and damage sites, A dual targeting that brings new insight.	Pt-ttpy covalently binds to G-quadruplex (G4) structures in vitro and to telomeres in cellulo via its Pt moiety. Here, we identified its targets in the human genome, in comparison to Pt-tpy, its derivative without G4 affinity, and cisplatin. Pt-ttpy, but not Pt-tpy, induces the release of the shelterin protein TRF2 from telomeres concomitantly to the formation of DNA damage foci at telomeres but also at other chromosomal locations. (Abstract)	A sonnet about a platinum complex named Pt-ttpy that binds G-quadruplex structures and induces telomere dysfunction.

After the extraction is complete, you may reprocess the prompts, change the prompts, delete/add PDFs, etc. any number of times as you desire.

You can also summarize the answer to any question/prompt across all your uploaded documents. To do so, click on the "Summarize" button on the top of the desired column:

Q2	Q3	Q4	
Summarize	Summarize	Summarize	
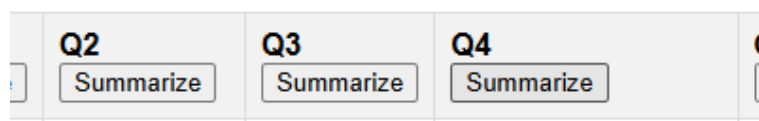
You may use the default summarization prompt provided, or customize as desired:

Summarize Responses for Q4 ✕

"What is the Research question or aim of the study?",

Summarization Prompt:

```
Summarize the responses provided as follows: Provide a detailed, step-by-step analysis of how you assessed the
recurrent themes represented in the responses, followed by how you categorized the responses by these themes.
Finally, provide a bulleted, point-by-point summary of these themes, an explanation of each theme, a count of the
number of responses pertaining to the theme, and representative quotes from the responses illustrating each theme.
```

Submit

Step-by-Step Analysis of Recurrent Themes

1. **Data Familiarization:**

- **Reading Responses:** Carefully reviewed each of the provided study aims to understand their focus and objectives.
- **Noting Initial Impressions:** Identified preliminary patterns and areas of overlap among the responses.

2. **Coding and Identification of Themes:**

ENJOY!