Organization Agreement
to use the

ClueWeb22 Web Research Collections

The _____ ("Group"), a group or division of approximately _____ people engaging in research and development of natural language processing, information-retrieval, or deep learning and/or related AI technologies, is part of the following corporation/partnership/legal entity listed below (the "Organization").

Corporation/Partnership/Legal Entity _____

Official mail address _____

_____

_____

_____

Contact for licensing:
Name:_____

Telephone_____

email :_____

Contact for technical matters/Information usage:
Name:_____

Telephone_____

email :_____

The Group would like to use the information designated as the ClueWeb22 Text Research Collections (the "Information"). By signing this Organization Agreement ("Agreement") with Carnegie Mellon University ("Carnegie Mellon"), the Organization hereby agrees to abide by the following understandings, terms and conditions. These understandings, terms and conditions apply equally to all or to part of the Information, including any updates or new versions of the Information supplied under this Agreement. Organization understands and agrees that the licenses to the Information granted under this Agreement are for use only by persons working within Organization's specific Group identified above, subject to the terms and conditions below.

 Initials

Copyright

1. The Information has been obtained by crawling the Internet and from publicly available web pages selected from the Bing search engine provided by Microsoft Corporation. Due to the amount of Information it has not been practicable to obtain permission from copyright owners to provide the Information for the uses permitted under this Agreement ("Permitted Uses").
2. Organization understands that all the documents in the Information are documents which have been at some time made publicly available on the Internet.
3. Owners of copyright in individual documents may choose to request deletion of these documents from the Information, and Carnegie Mellon may require deletion of certain documents from the Information.
4. The limitation on permitted use contained in the following section is intended to reduce the risk of any action being brought by copyright owners, but if this happens the Organization agrees to bear all associated liability.

Permitted Uses

1. The Information may only be used for research and development of information retrieval, natural language processing, deep learning and/or related AI technologies purposes by the specific Group identified above.
2. Summaries, analyses and interpretations of the linguistic properties of the Information may be derived and published, provided it is not possible by the Organization or anyone else using such summaries, analyses and interpretation to reconstruct the Information from these summaries.
3. In addition, small excerpts of the Information may be displayed to others or published in a scientific or technical context, solely for the purpose of describing the research and development carried out and related issues, provided it is not possible by the Organization or anyone else using such excerpts to reconstruct the Information from these excerpts.
4. All efforts must be made not to infringe the rights of any third party including, but not limited to, the authors and publishers of any excerpts used in accordance with clause 3 above in this "Permitted Uses" section.
5. The Information shall be used by Organization in compliance with all applicable laws, rules, and regulations.

Own Assessment of Information Use

The Organization must make its own assessment of the suitability of the Information for its research and development purposes under Permitted Uses and that Organization's use of the Information is consistent with applicable laws, rules and regulations.

Agreement to Delete Data on Request

The Organization shall immediately delete upon receiving notice all copies of any particular document that is part of the Information whenever requested to do so by either:

1. Carnegie Mellon; or
2. the owner of copyright for the particular document.

Access to the Information by Individuals

The Organization:

1. must control access to the Information by individuals and may only grant access to people within the identified Group who are working under its control, i.e., its own employees, consultants under written agreement to the Organization, or individuals providing service to the Organization under written agreement;
2. must ensure that before being given access an individual must complete and submit the Individual Agreement form as attached hereto;
3. must terminate an Individual's access when the individual no longer requires access for its work for the Organization and/or no longer is employed by (and/or under contract with, as applicable) the Organization;
4. remains responsible for any breach of the Individual Agreement form by individuals to whom Organization has granted access to the Information;
5. shall retain the applications of all persons ever granted access to the Information and make them available upon request to any of the copyright holders and to Carnegie Mellon;
6. shall maintain a list of people with current and recently-terminated access to the Information and make it available to Carnegie Mellon on request; and
7. must make sure that an Individual with access displays the Information to or shares the Information with only persons whom his or her Organization lists as having access to the Information.

No Warranty; Disclaimers; Indemnification

THE INFORMATION IS PROVIDED ON AN "AS IS" BASIS. NO REPRESENTATION OR WARRANTY OR CONDITIONS OF ANY KIND ARE GIVEN BY CARNEGIE MELLON, MICROSOFT CORPORATION, OR ANY SOURCE FROM WHICH MICROSOFT CORPORATION DIRECTLY OR INDIRECTLY RECEIVED MATERIAL THAT IS INCLUDED IN THE INFORMATION (EACH SUCH SOURCE BEING AN "UPSTREAM DATA PROVIDER") IN RELATION TO THE INFORMATION OR THE USE(S) TO WHICH THE INFORMATION MAY BE PUT BY THE ORGANIZATION, INCLUDING WITHOUT LIMITATION ANY REPRESENTATIONS OR WARRANTIES THAT THEY HAVE ANY RIGHTS IN THE INFORMATION OR ANY CONDITIONS OF TITLE, NON-INFRINGEMENT, MERCHANTABILITY OR THE FITNESS OR SUITABILITY OF THE INFORMATION FOR ANY PARTICULAR PURPOSE OR UNDER ANY SPECIAL CONDITIONS REGARDLESS OF WHETHER ANY SUCH PARTICULAR PURPOSE OR SPECIAL CONDITIONS ARE OR HAVE BEEN MADE KNOWN TO CARNEGIE MELLON, MICROSOFT CORPORATION, OR ANY

UPSTREAM PROVIDER PRIOR TO, OR DURING, THE PERIOD OF THIS AGREEMENT. ANY AND ALL SUCH CONDITIONS AND WARRANTIES EXPRESS OR IMPLIED, WHETHER ARISING UNDER STATUTE OR UNDER COMMON LAW, INCLUDING BUT NOT LIMITED TO CONDITIONS AND WARRANTIES AS TO QUALITY, MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NON-INFRINGEMENT ARE HEREBY EXCLUDED AND EXPRESSLY DISCLAIMED. NEITHER CARNEGIE MELLON, MICROSOFT CORPORATION, NOR ANY UPSTREAM PROVIDER SHALL BE LIABLE TO ORGANIZATION OR ANY THIRD PARTY FOR LOSS OF PROFITS OR FOR INCIDENTAL, INDIRECT, SPECIAL OR CONSEQUENTIAL DAMAGES FOR ANY REASON WHATSOVER ARISING OUT OF OR RELATING TO THIS AGREEMENT (INCLUDING ANY BREACH OF THIS AGREEMENT), EVEN IF SUCH PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES OR HAS OR GAINS KNOWLEDGE OF THE EXISTENCE OF SUCH DAMAGES. ORGANIZATION FURTHER ACKNOWLEDGES AND AGREES THAT THIS AGREEMENT IS BETWEEN ORGANIZATION AND CARNEGIE MELLON, AND TO THE MAXIMUM EXTENT PERMITTED UNDER APPLICABLE LAW NEITHER MICROSOFT NOR ANY UPSTREAM PROVIDER SHALL HAVE LIABILITY FOR ANY DIRECT DAMAGES UNDER THIS AGREEMENT.

ORGANIZATION SHALL DEFEND, INDEMNIFY AND HOLD HARMLESS CARNEGIE MELLON AND MICROSOFT CORPORATION AND EACH OF THEIR RESPECTIVE DIRECTORS/TRUSTEES, OFFICERS, EMPLOYEES, ATTORNEYS AND AGENTS FROM AND AGAINST ANY LIABILITY, DAMAGE, LOSS OR EXPENSE (INCLUDING ATTORNEYS' FEES AND EXPENSES) INCURRED BY OR IMPOSED UPON ANY OF CARNEGIE MELLON AND MICROSOFT CORPORATION AND/OR ANY OF THEIR RESPECTIVE DIRECTORS/TRUSTEES, OFFICERS, EMPLOYEES, ATTORNEYS AND AGENTS IN CONNECTION WITH ANY CLAIM, SUIT, ACTION OR DEMAND ARISING OUT OF OR RELATING TO ANY EXERCISE OF ANY RIGHT OR LICENSE GRANTED OR PROVIDED TO ORGANIZATION OR ANY FAILURE TO PERFORM ANY OBLIGATION OF ORGANIZATION UNDER THIS AGREEMENT UNDER ANY THEORY OF LIABILITY (INCLUDING WITHOUT LIMITATION, ACTIONS IN THE FORM OF TORT, WARRANTY, OR STRICT LIABILITY, OR VIOLATION OF ANY LAW, AND REGARDLESS OF WHETHER SUCH ACTION HAS ANY FACTUAL BASIS).

Termination

Either party may terminate this Agreement at any time by notifying the other party in writing. On termination, the Organization must: a) immediately cease using the Information; and b) delete all copies of the Information.

Applicable Law; Disputes

This Agreement is governed by the laws of the Commonwealth of Pennsylvania in the United States of America. All claims and/or controversies of every kind and nature arising out of or relating to this Agreement, including any questions concerning its existence, negotiation, validity, meaning, performance, non-performance, breach, continuance or termination shall be settled exclusively in the United States District Court for the Western District of

Pennsylvania or, if such Court does not have jurisdiction, in any court of general jurisdiction in Allegheny County, Pennsylvania and each party consents to the exclusive jurisdiction of any such courts and waives any objection which such party may have to the laying of venue in any such courts.

Notices

Notices to the Organization may be provided either by electronic or physical mail to the licensing contact listed on the first page of this Agreement. Notices to Carnegie Mellon may be provided in the same manner to the following:

Director of Technology Licensing
Center for Technology Transfer and Enterprise Creation
Carnegie Mellon University
4615 Forbes Avenue
Pittsburgh, PA 15213
USA

telephone: +1 412-268-7393
email: innovation@cmu.edu

Either party may update its contact information by providing written notice to the other party as required by this Section.

Third Party Beneficiaries

Microsoft Corporation is an express and intended third party beneficiary of this Agreement and shall have the right to independently enforce the terms of this Agreement against Organization as if Microsoft was a party to this Agreement. Except as provided in the prior sentence, there are no other third party beneficiaries of this Agreement and only Carnegie Mellon and Organization shall be entitled to enforce any rights, benefits or remedies pursuant to this Agreement.

Miscellaneous

If any portion of this Agreement is determined by any court or governmental agency of competent jurisdiction to violate applicable law or otherwise not to conform to requirements of law, then the rest of the Agreement will remain in effect and the parties will substitute a suitable and equitable provision for the invalid/unenforceable provision in order to carry out the original intent and purpose of the original Agreement. Organization may not assign any or all of its rights and/or obligations under this Agreement without the prior written consent of Carnegie Mellon, which consent may be granted or withheld in Carnegie Mellon's sole discretion. Any attempted assignment in violation of this section shall be void and of no effect. This Agreement constitutes the entire agreement between the parties and supersedes all previous agreements and understandings relating to the subject matter of this Agreement. The Agreement may not be altered, amended, or modified except by a written instrument signed by the duly authorized representatives of both parties.

Intending to be legally bound, Organization and Carnegie Mellon execute this Agreement effective as of the date the last party signs.

By the Organization:

By signing below, I represent and warrant that I have authority to bind the Organization to the terms of this Agreement

Signature_____

Date _____

Name (please print) _____

Title _____

Accepted by Carnegie Mellon University:

Signature _____

Date _____

Name (please print) _____

Title _____

ClueWeb22 Order Form

Which version(s) of the dataset do you want? (Most people pick just one)

Document Category	Document Format(s)	Distribution Media	Cost	
All	All	Dataset license only	$0	☐
B	txt	Download (511 GB)	$0	☐
B	txt	DeepResearchGym API	$0	☐
B	txt	1 × 1 TB disk	$310	☐
B	html, txt, links, vdom	1 × 18 TB disk	$715	☐
B	jpg	6 × 18 TB disk	$3,870	☐
A (includes B)	html, txt, links, vdom	8 × 18 TB disk	$4,985	☐
L	txt, links	2 × 18 TB & 1 × 8 TB disk	$1,530	☐
TREC-iKAT-2023	txt	Download (26 GB)	$0	☐
TREC-LR-2024-T1	txt	Download (0.5 MB)	$0	☐

If the dataset will be shipped to you on disk(s):

1. **Where should the invoice be sent?**

   Name: _____

   Email: _____

2. **Where should the dataset be sent?**

   Name: _____

   Email: _____

   Mailing address: _____

   _____

   _____

   _____

   _____

   Telephone: _____

3. **Preferred shipping method:** Standard ☐ Priority ☐

If the dataset will be downloaded or accessed through the DeepResearchGym API:

Name: _____

Email: _____