REPORT DOCUMENTATION PAGE	Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
12-12-2022	Final Report	1-Oct-2018 - 30-Sep-2022

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Final Report: Collaborative World Modeling	W911NF-18-1-0464
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHORS	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES	8. PERFORMING ORGANIZATION REPORT NUMBER
Florida Institute for Human and Machine Cc 40 S. Alcaniz St. Pensacola, FL 32502 -6008	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	ARO
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) 74058-MI-DRP.1

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			James Allen
UU	UU	UU	UU		19b. TELEPHONE NUMBER 850-202-4473

Standard Form 298 (Rev 8/98)
Prescribed by ANSI Std. Z39.18

RPPR Final Report
as of 13-Dec-2022

Agency Code: 21XD

Proposal Number: 74058MIDRP
Agreement Number: W911NF-18-1-0464

INVESTIGATOR(S):

Name: James Allen
Email: jallen@ihmc.us
Phone Number: 8502024473
Principal: Y

Organization: **Florida Institute for Human and Machine Cognition Inc.**
Address: 40 S. Alcaniz St., Pensacola, FL 325026008
Country: USA
DUNS Number: 158995659
EIN:

Report Date: 31-Dec-2022
Date Received: 12-Dec-2022
Final Report for Period Beginning 01-Oct-2018 and Ending 30-Sep-2022
Title: Collaborative World Modeling
Begin Performance Period: 01-Oct-2018
End Performance Period: 30-Sep-2022
Report Term: 0-Other

Submitted By: Choh Man Teng
Email: cmteng@ihmc.us
Phone: (850) 202-4469

Distribution Statement: 1-Approved for public release; distribution is unlimited.

STEM Degrees: 0
STEM Participants:

Major Goals: The current state of the art in world modeling involves such manual effort that it is infeasible to build large models fast enough to enhance short-term decision making about issues of national and global security. We propose an effort to automate large parts of the building and running of such models, enabling an order of magnitude reduction in development time and a significant increase in the scale of problems that can be explored. The key insight is that developing large-scale world models requires a collaboration between human analysts (providing intuition, strategic thinking and thinking out of the box) and automated intelligent systems (providing large scale data search and analysis, planning and running simulations). By harnessing human-machine collaboration, we enable rapid development of models that go far beyond what humans or automated methods could each achieve alone. The key to success is that model development is iterative, with extensive interaction between the analyst and the system throughout.

We will leverage our existing technology base of (1) collaborative problem solving through human-machine interaction, (2) planning and (3) task learning, to build an open, extensible and interactive system for World Modeling. Quantitative Reasoning Engines (QREs) can be plugged in as they become available (or taken offline), using a uniform messaging structure for declaring their inputs, outputs and data requirements. The parameters will be encoded in a standardized ontology, and matched with natural language text that might describe these parameters using a module that scores the semantic similarity between the parameter documentation and the natural language text. The system will use this information, together with information obtained from reading systems or from interaction with Subject Matter Experts, to build and continually update its causal models. It will work with subject matter experts to exploit available Quantitative Reasoning Engines, and identify gaps in the workflow and workarounds.

Accomplishments: See attached.

Training Opportunities: Nothing to Report

Results Dissemination: Nothing to Report

Honors and Awards: Nothing to Report

Protocol Activity Status:

Technology Transfer: Nothing to Report

PARTICIPANTS:

Participant Type: Other Professional
Participant: William de Beaumont
Person Months Worked: 14.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: PD/PI
Participant: James Allen
Person Months Worked: 10.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Staff Scientist (doctoral level)
Participant: Archna Bhatia
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Staff Scientist (doctoral level)
Participant: Lucian Galescu
Person Months Worked: 12.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Staff Scientist (doctoral level)
Participant: Choh Man Teng
Person Months Worked: 12.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Partners

,

I certify that the information in the report is complete and accurate:
Signature: Choh Man Teng
Signature Date: 12/12/22 5:29PM

Collaborative World Modeling

World Modelers

Final Report (2018/10-2021/02)

Summary

The current state of the art in world modeling involves such manual effort that it is infeasible to build large models fast enough to enhance short-term decision making about issues of national and global security. We propose an effort to automate large parts of the building and running of such models, enabling an order of magnitude reduction in development time and a significant increase in the scale of problems that can be explored. The key insight is that developing large-scale world models requires a collaboration between human analysts (providing intuition, strategic thinking and thinking out of the box) and automated intelligent systems (providing large scale data search and analysis, planning and running simulations). Model development is iterative, with extensive interaction between the analyst and the system throughout. The key components of our solution are: 1) an intuitive interface for iterative model development and evaluation, using a language-driven Collaborative Problem Solving process between users and World Modeling Systems; 2) automatic and human-assisted transformation of qualitative models to executable quantitative models by means of collaborative planning; 3) capabilities for users to define novel problems, bootstrapping from a domain-independent ontology; 4) a system that learns and improves as it interacts with its users.

1. Text Reading

One of the key points for better reading is that if we are to use reading to build models of specific situations, the reading needs to be able to extract a significant amount of detailed information from a relatively small set of key documents (targeted deep reading), rather than small amounts of information from a multitude of papers (shallow reading as in the current reading efforts to build causal graphs). Specifically, constructing a model of situations involves identifying spatially and temporally grounded events, with their full complement of arguments. Many of these events are not causal relations, although of course extracting the described causal relations between the events is equally important.

The TRIPS reading system is ideally suited for this more detailed reading as it produces a fairly-detailed representation of the document contents which is then simplified by extracting the information of relevance. We developed CWMSReader, a paper reading system that parses documents in the World Modelers domain. CWMSReader parses sentences into Logical Forms, which expresses the meaning of a sentence in terms of the TRIPS ontology. Concepts and relations of particular interest are further extracted and normalized. The target extractions can be domain-independent, for example causal relations, or they can be domain-specific, for example interventions.

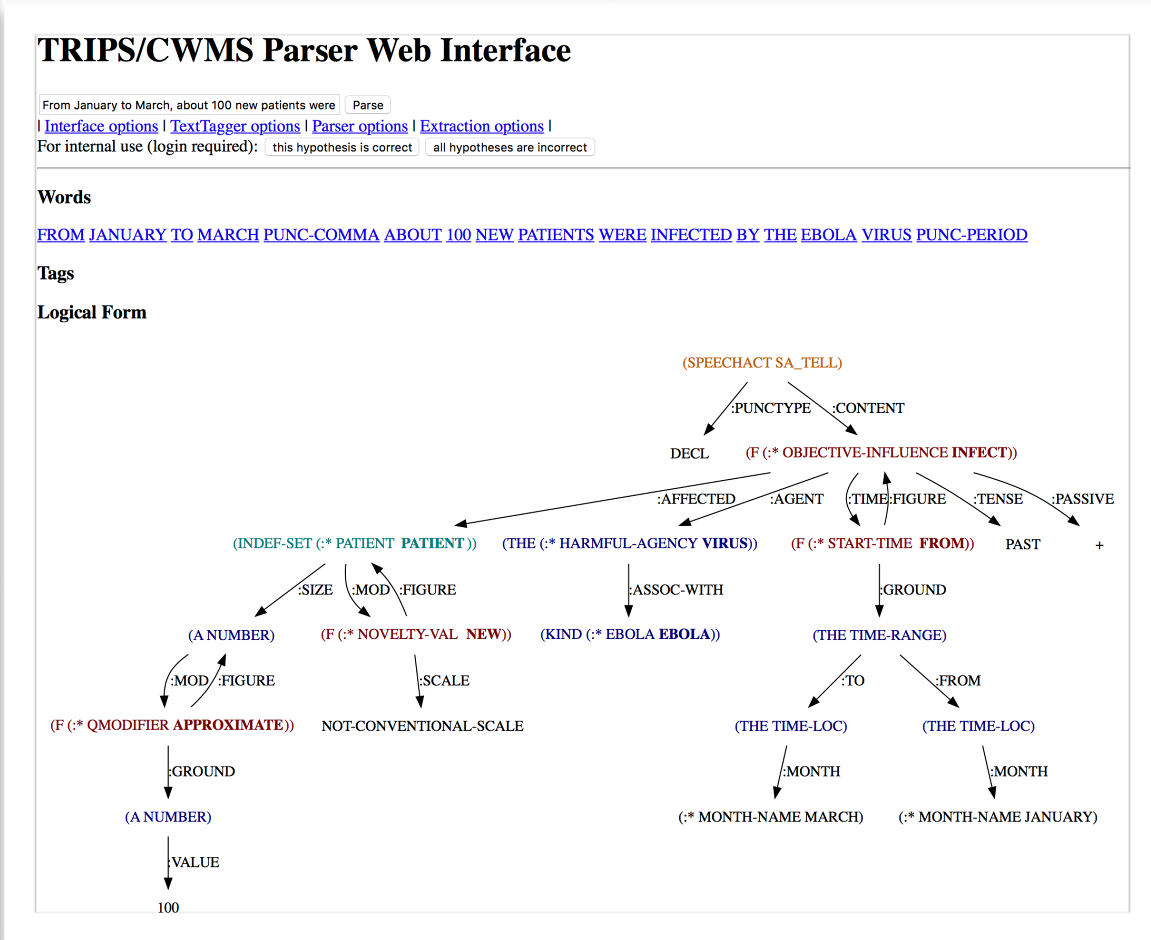

Figure 1. Parsing temporal and numeric expressions

To support extraction of relevant information, we improved parsing in several fundamental areas, including temporal and number expressions, especially ranges (e.g., June to October, 11/12/08-01/31/09, five to seven, …); unit expressions, including standard rates (e.g., miles per hour) and idiosyncratic rates (e.g., seeds per acre), and locations. CWMSReader is available as a web service with associated API at http://trips.ihmc.us/parser/cgi/cwmsreader. Figure 1 shows an example of parsing a sentence with date ranges and approximate numeric expressions, using the CWMSReader web interface. Figure 2 shows a parse of a sentence containing several intervention events.

In addition, we improved the handling of words not explicitly defined in the TRIPS lexicon by incorporating entries derived from WordNet, an external dictionary. This enabled better semantic analysis of 100,000+ words and enhanced the domain-independent generation of semantics with the broad coverage of English. The extended TRIPS ontology and lexicon browser is available at http://trips.ihmc.us/parser/cgi/lex-ont (Figure 3).

We identified and incorporated terms that are domain-specific and are not expected to be found in general English dictionaries. These include *meher*, *belg* and others for various rainy and growing seasons; and *woreda* and *kebele* for administrative divisions. We also developed

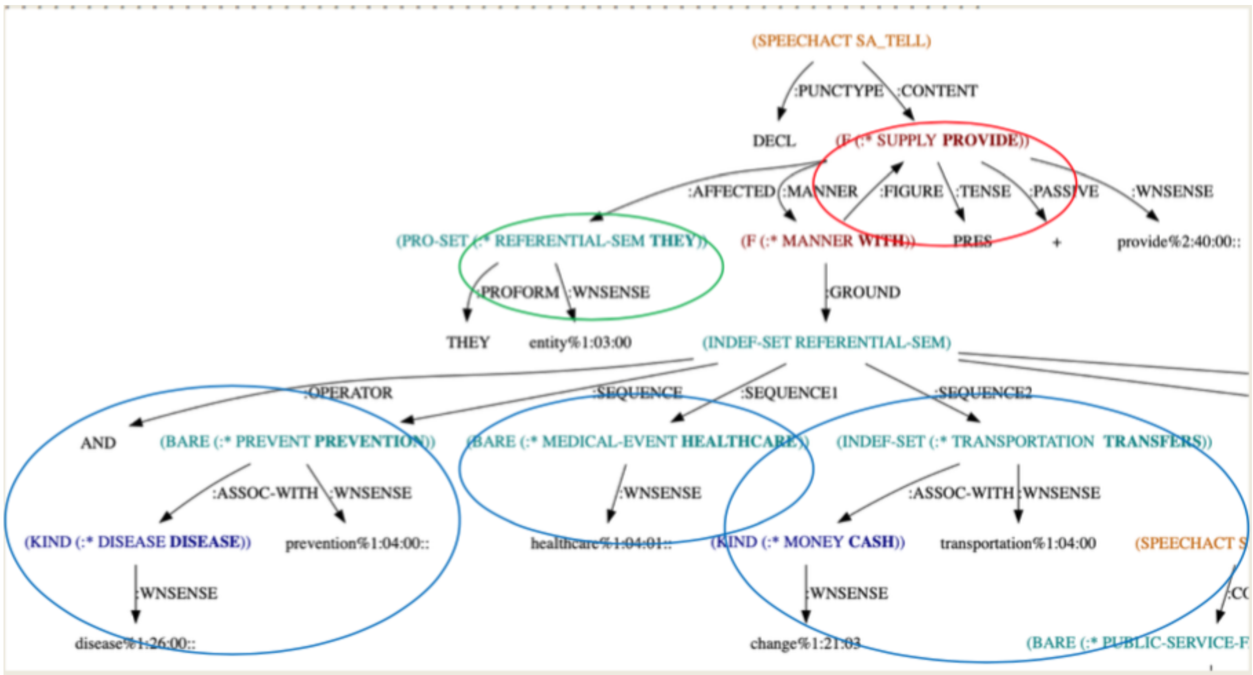

Figure 2. **They** are **provided** with safe water, shelter, food aid, basic essential items, **cash transfers**, **healthcare**, and **disease prevention**.

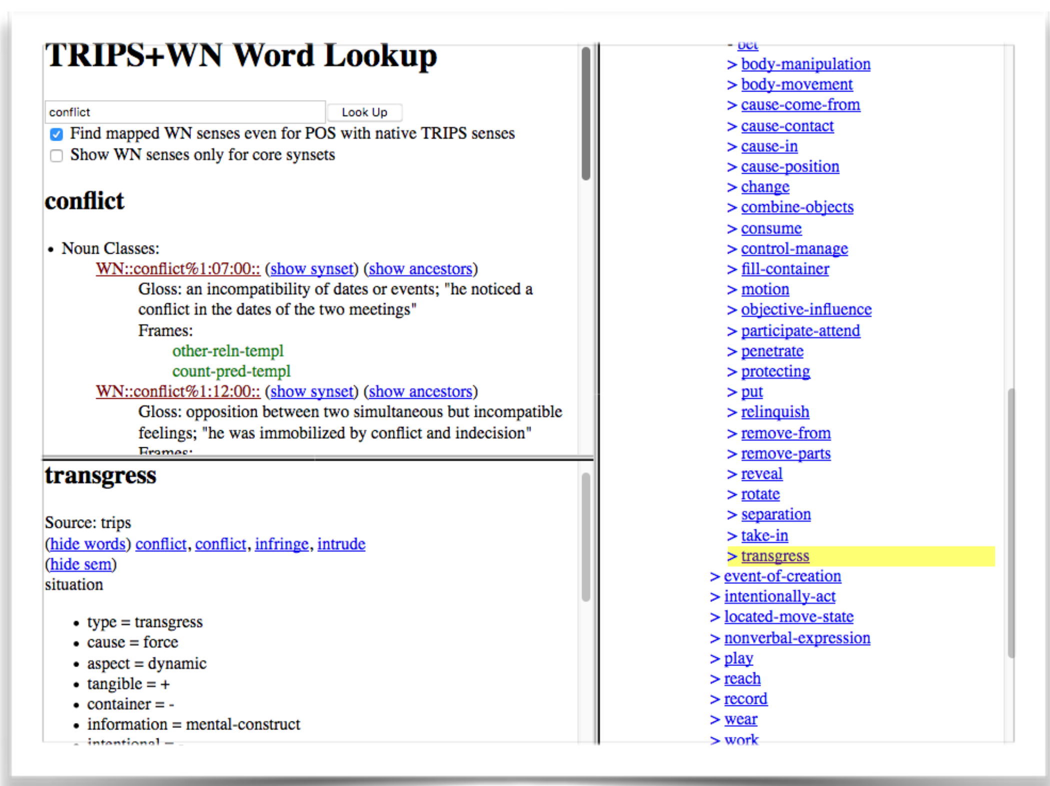

Figure 3. Extended TRIPS ontology and lexicon browser

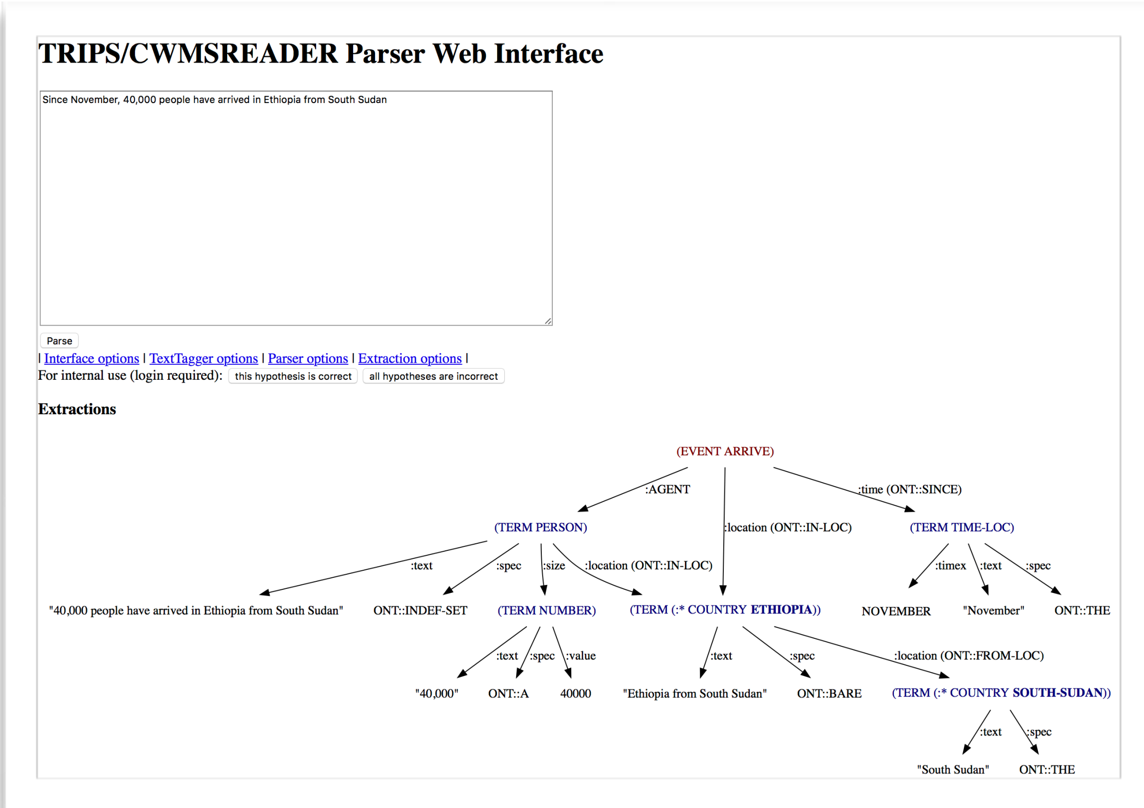

Figure 4. Extraction of a migration event

mappings from the TRIPS ontology to the World Modelers intervention ontology.

2. EKB Extractions

To facilitate downstream processing, relevant information is extracted and standardized into an XML-based format called the EKB, for input to INDRA. In addition to domain independent extractions such as causal information, we focused on several areas of particular interest, such as migration and agriculture. Figure 4 shows an example of the extraction of a migration event. Figure 5 shows multiple causal extractions from a sentence. The EKB representation has been extended to extract domain-focused information, as well as time and location information, in a normalized and simplified format for ingestion by INDRA. The EKB extractions include annotation fields to encode descriptions of the underlying data, such as provenance and the time of processing.

Below are a few other examples that CWMSReader can extract from the text.

- In Ethiopia, **massive deforestation** of natural forests and extensive use of agricultural lands **have resulted in soil degradation**.
- […] **rainfall and days to maturity have a large positive influence on yield**
- […] **the mass exodus of people has resulted in reduction of productivity**

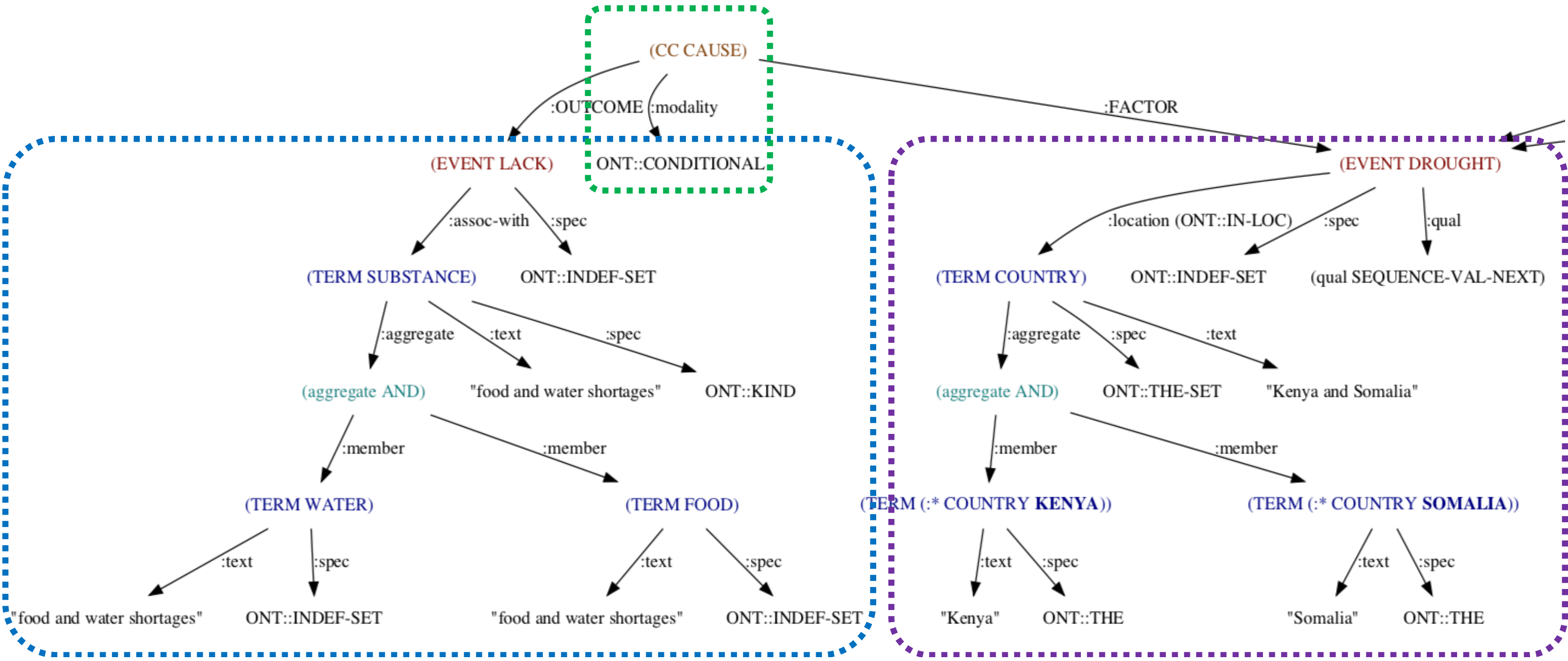

Figure 5. A causal extraction from the sentence "*Back-to-back **droughts in Kenya and Somalia**, or a March-June drought in Ethiopia, **could cause food and water shortages**, hurt hydro-power stations and worsen farming conditions.*" Not shown are additional extractions for "[droughts] *hurt hydro-power stations*" and "[droughts] *worsen farming conditions*".

- In Somali region, **the prolonged drought has led to the displacement of many children** who remain out of school
- […] **producing more crops will lead to improving access to these foods by rural people** with limited resources.

3. Unstructured Data Pipeline

CWMSReader was integrated with DART /Kafka. We provided feedback to TwoSix on the new CDR format and the DART-readers integration process. We implemented a Kafka stream consumer for CDRs produced by DART, and an EKB output stream uploader to transfer reader output to storage. We updated the CWMSReader workflow for batch processing. Figure 6 shows the DART-CWMSReader workflow.

In preparation of large scale reading experiments, we made updates to CWMSReader, in particular with respect to the management of very large documents and the speed of processing. We also improved the handling of failures and developed heuristics for distinguishing language and non-language sections in input documents.

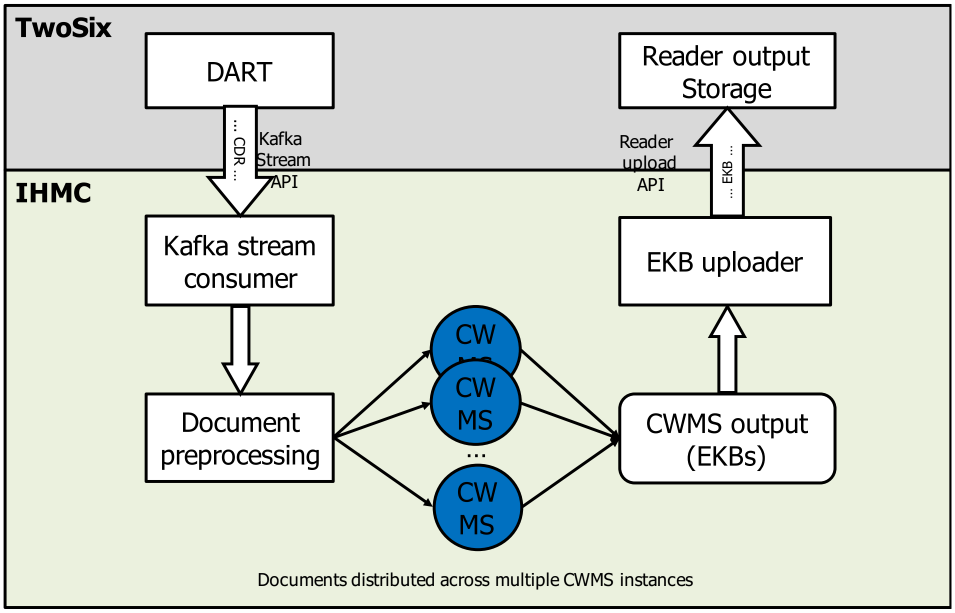

Figure 6. The DART-CWMSReader workflow.

Temporal expressions, not only standalone dates but also ranges and durations (e.g., March through June, first half of 1998, the end of April) are extracted in a normalized form. Vague and approximate numeric expressions, e.g., over/around/almost 2300, are handled appropriately. In addition, we improved the reading of both text and tables to extract domain relevant information.

To facilitate the evaluation of changes to our natural language understanding components, we collected representative sentence samples from the input documents used in the reading experiments and elsewhere and set up a test harness that compares and scores parses and extractions with respect to stored baselines of these sentences. This provides validation of improvements and monitors the outputs such that unintentional changes can be detected in a timely manner.

We also developed the capability to generate EKB files from the output of table reading, providing a uniform interface to INDRA for text and table reading.

4. Table Reading

Documents such as reports, news articles and scientific papers often contain tables. The text narrative refers to the information laid out in the tables but typically does not repeat it in detail. On the other hand, the captions and headings of a table are often abbreviated and can only be fully interpreted in the context of the text narrative. We developed a a module for reading tables in PDF documents automatically and extracting the information into the EKB format used also for the extraction from text.

The table reader is capable of detecting the location of a table in a document, identifying the layout of the table and extracting the values of the cells, suitably aligned into a spreadsheet-

Tifton, Georgia:
Soybean Variety Performance, 2017, Irrigated

Company or Brand Name	Variety	2-Year Average Yield	2017 Data						
			Rank	Yield[1]	Maturity	Plant Ht	Lodg.[2]	Wt of 100 Seed	Seed Quality[3]
		bu/acre		bu/acre	date	in	rating	gm	rating
Maturity Group V									
Dyna-Gro	39RY57	**74.5**	4	**66.0**	09/26	34	4.0	16.0	1.5
Bayer	CZ 5147 LL	**63.7**	5	**64.7**	09/23	27	2.7	14.8	1.5
Virginia Tech	V12-1416	**63.3**	6	**64.3**	10/04	31	2.3	14.9	1.5
UARK	Osage	**62.8**	13	60.9	09/28	27	3.7	13.3	1.5
Meherrin	SH 5215 LL	61.8	8	62.4	09/18	39	4.0	13.5	1.5
Meherrin	SH 5915 LL	60.1	31	53.0	09/27	35	4.3	15.3	1.5
Bayer	CZ 5242 LL	59.8	16	58.6	09/19	41	5.0	12.4	1.5
UARK	UA 5414RR	59.2	22	57.0	09/29	33	4.7	14.5	2.3
Syngenta	S58-Z4	58.5	14	59.9	10/03	32	4.0	13.9	1.5
Bayer	CZ 5375 RY	57.4	35	51.6	09/27	29	3.7	13.9	1.5
USDA-ARS	JTN-5110	53.4	40	48.0	09/29	32	4.0	16.6	1.7
Bayer	CZ 5515 LL	53.4	43	40.0	09/27	43	5.0	14.1	1.5
Dupont Pioneer	P55A49X	.	1	**72.8**	09/27	29	1.7	13.9	1.7
Virginia Tech	V13-3833	.	2	**70.5**	09/23	29	1.3	15.1	1.7
Armor	47-D17	.	3	**66.4**	09/26	37	4.0	13.8	2.2
Virginia Tech	V12-0045 R2	.	7	64.0	09/27	29	2.7	16.7	1.8
UARK	UA 5014C	.	9	62.1	09/22	29	3.3	17.1	1.7
[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Figure 7. A table with abbreviations, footnotes and headings spanning multiple rows and columns

compatible representation according to the layout. Figure 7 shows a typical table. The PDF format gives rise to a number of complications. For example, footnote notations may not be explicitly marked but follow directly the value in a cell, and headings as well as cells sometimes graphically span multiple rows and columns. The table reader is able to meet reasonably well many of these challenges, including

- detecting paragraph regions and fixing the location of a table
- extracting functional components and properties such as row and column headings, footnotes and captions
- splitting and merging of cells, rows and columns
- processing footnotes
- handling of symbols and other artifacts from PDF-to-text conversions
- resizing the window when the page size changes
- undo and redo operations

The texts in tables, often short and abbreviated phrases or incomplete sentences, are crucial in the interpretation of the table cell values. These texts need to be parsed into a normalized form and used for further analysis of the table data.

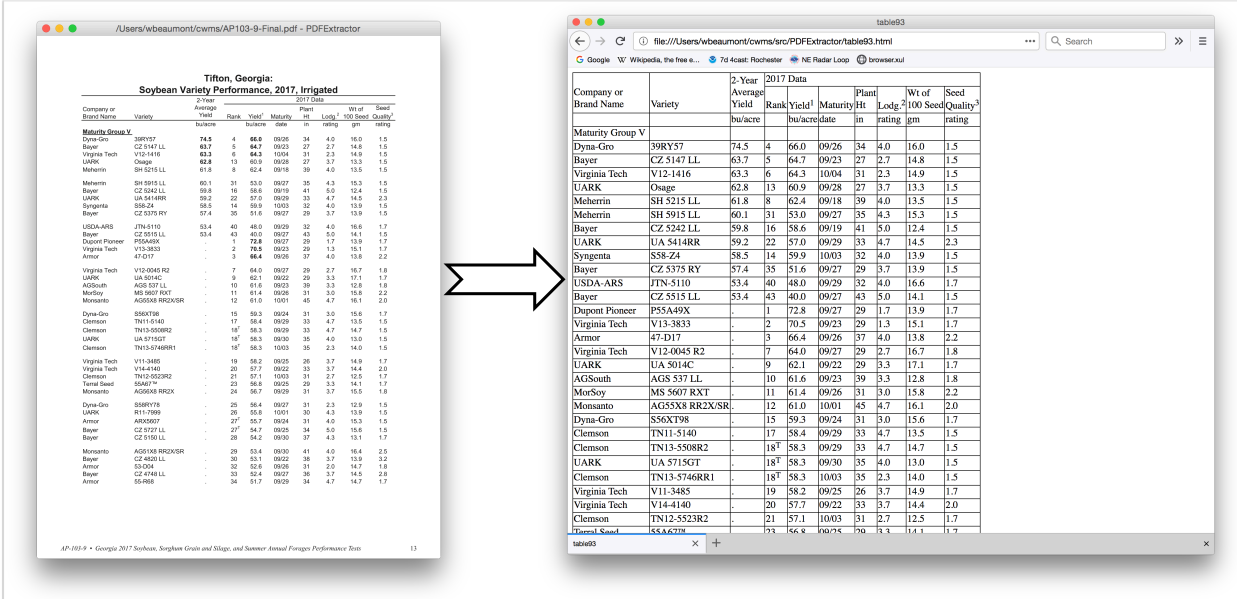

Figure 8. Data extracted from PDF tables according to the table layout are converted into spreadsheet-like format that supports interactive correction and modification by a user

Figure 8 shows an example of extraction from the table in Figure 7. This use case was focused on auto-calibrating the genetic coefficients for a crop/cultivar by reading tables that contained variety trials data.

The table reader has been incorporated as a component in the CWMS dialogue system to support human-user collaborative interpretation of tables. Through interactive editing, the user can guide the system and correct the system's errors in identifying the layout of the table.

In parallel we developed a capability for the table reader to learn from this interaction to better identify tables in PDF documents, and to utilize appropriate editing operations to identify and align the table components in different styles of table encodings. Based on user input on identifying and segmenting a table, the system learns generalized rules that can be applied to handle other similar tables automatically.

The table reader is able to extract information from tables in a range of styles. Figure 9 shows the varied table styles the module can cope with. In particular, we also extended the table reading module to extract information from pseudo-tabular charts. Some examples are shown in Figure 10.

We created a combined text and table reading capability with multiple output modalities: the EKB format for input to INDRA, and the CSV format for spreadsheets. Figure 11 illustrates the integrated processing framework. This would enable the aggregation and cross-referencing of the information from the text and the information from the tables in a consistent manner.

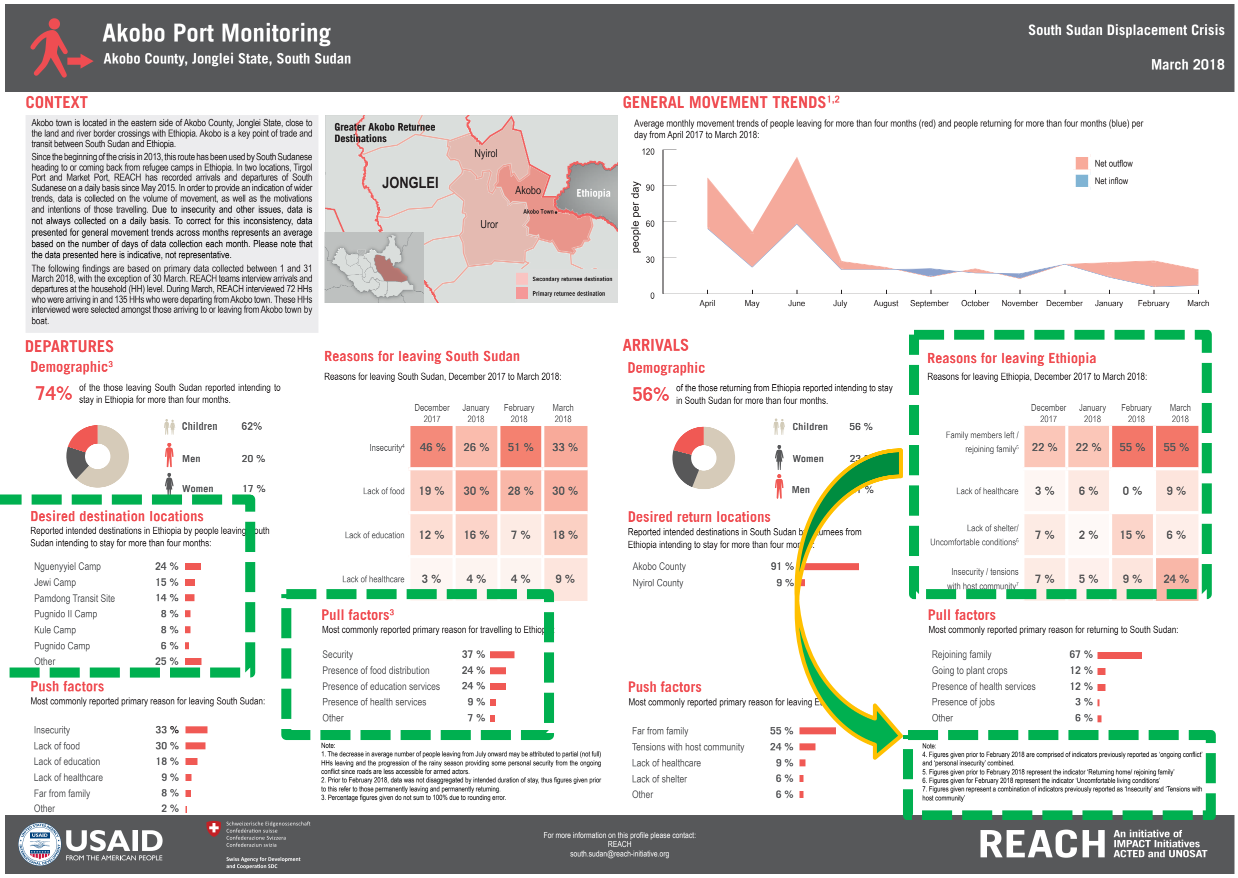

Figure 9. A variety of table styles handled by the table reading module. Examples are highlighted in green. Note that footnotes in a table might be found in a detached location in the document.

We demonstrated an end-to-end system in CWMS that can start with a user-selected table in a PDF document and extract information into the EKB format also used for extractions from text. The integrated extraction format facilitates uniform processing downstream in INDRA.

The table reading module has been refactored into a standalone application, based on feedback from Jataware and other collaborators, with extended documentation and examples. The table reading results can be output in a number of formats suitable for input into, for example, Excel.

5. Collaborative Problem Solving

Building models in a fully automatic fashion is not a realistic technical goal for a number of reasons. Even if read perfectly (an unrealistic assumption!), a set of documents is unlikely to describe everything required to build and instantiate a model. This is especially the case since much of the required detail is not to be found in the documents in the first place, as these details, though not mentioned explicitly, are obvious to any human reader using their commonsense

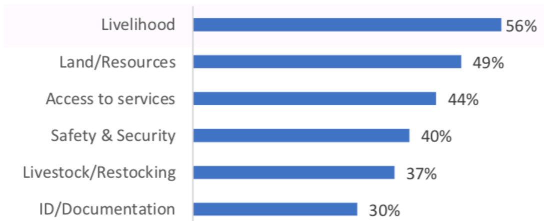

Figure 4: % of sites by support needed for return or reintegration

Livelihood	56%
Land/Resources	49%
Access to services	44%
Safety & Security	40%
Livestock/Restocking	37%
ID/Documentation	30%

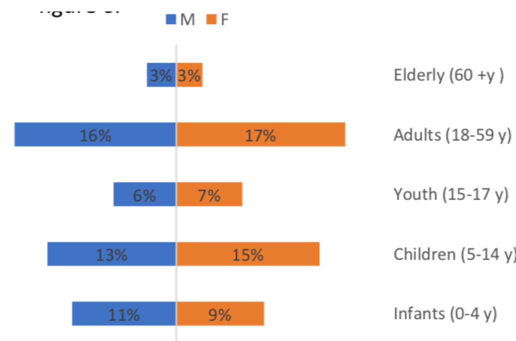

Figure 5: Gender and age distribution

3%	3%	Elderly (60 +y)
16%	17%	Adults (18-59 y)
6%	7%	Youth (15-17 y)
13%	15%	Children (5-14 y)
11%	9%	Infants (0-4 y)

Figure 10. Extraction of information from pseudo-tabular charts. Examples taken from *DTM_Round 8 Oromia region.pdf.*

knowledge. We believe that such models can be constructed effectively and efficiently by utilizing human-computer collaboration. Essentially, the machine builds partial models based on the read documents, and the human can identify and add critical implicit details in order to complete the model. Not only does this produce better models, but each time there is an interaction, the system can also learn some commonsense reasoning patterns that can be reused to improve subsequent reading tasks.

Equally important, the same collaborative interface can be used for the user to model "what-if" scenarios critical for intervention planning. Much of this information will not be found in documents and can only come from human experts. We believe this is probably the only viable mechanism for moving from models developed from reading into executable quantitative models that can be simulated to obtain actionable results.

The infrastructure for collaborative problem solving can be used to support interactive modeling, and exploration of what-if scenarios and intervention strategies. To enable collaborative CAG construction, we developed an EKB store as well as a graphical display and editing tool for EKBs. Events and relations extracted by CWMSReader can be selectively displayed based on the user's (natural language or GUI) commands. The user can query the current CAG for background information, and amend or correct the graph as needed.

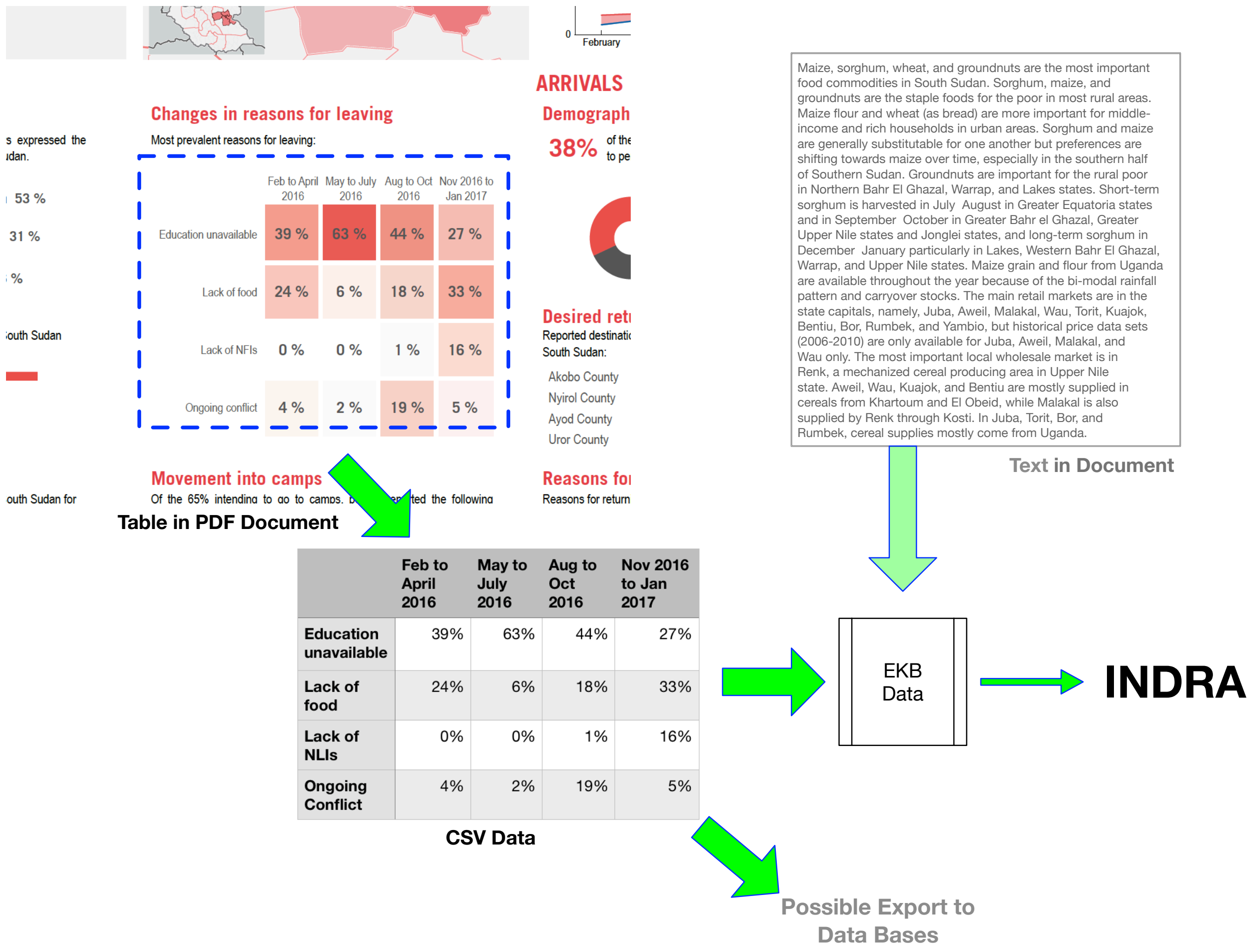

Figure 11. Integrating table reading and text reading

The collaborative human-machine interactive task specification handles the modelling of the ongoing problem solving process that the user is engaged in, enables effective intention recognition and facilitates proactive system behavior. The framework supports natural language interaction in all circumstances. This task model enables a system that proactively participates in solving problems, going well beyond responding literally to user commands.

We completed the design and implementation of a new Collaborative Behavioral Agent to support interaction with the user. This capability is key to providing intuitive user access to complex black box reasoners and enhances user confidence in system results. Figure 12 shows the Behavioral Agent in the context of World Modelling. Figure 13 shows the schematics of the CWMS Collaborative Problem Solving Model.

Together with researchers at University of Florida, we developed several scenarios of collaborative modeling, using DSSAT as the underlying modeling engine. The human and machine engage in a dialogue to incrementally build a model, run simulations and explore effects of interventions. The information needed was gleaned from both the text narratives and the tables found in documents. Figure 14 shows the processing of a table interactively. Figure 15

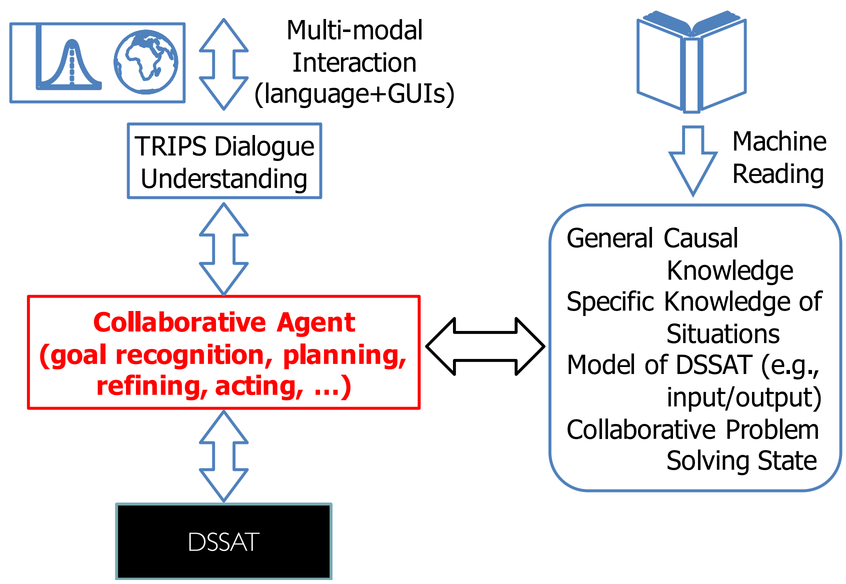

Figure 12. Behavioral Agent for World Modelling

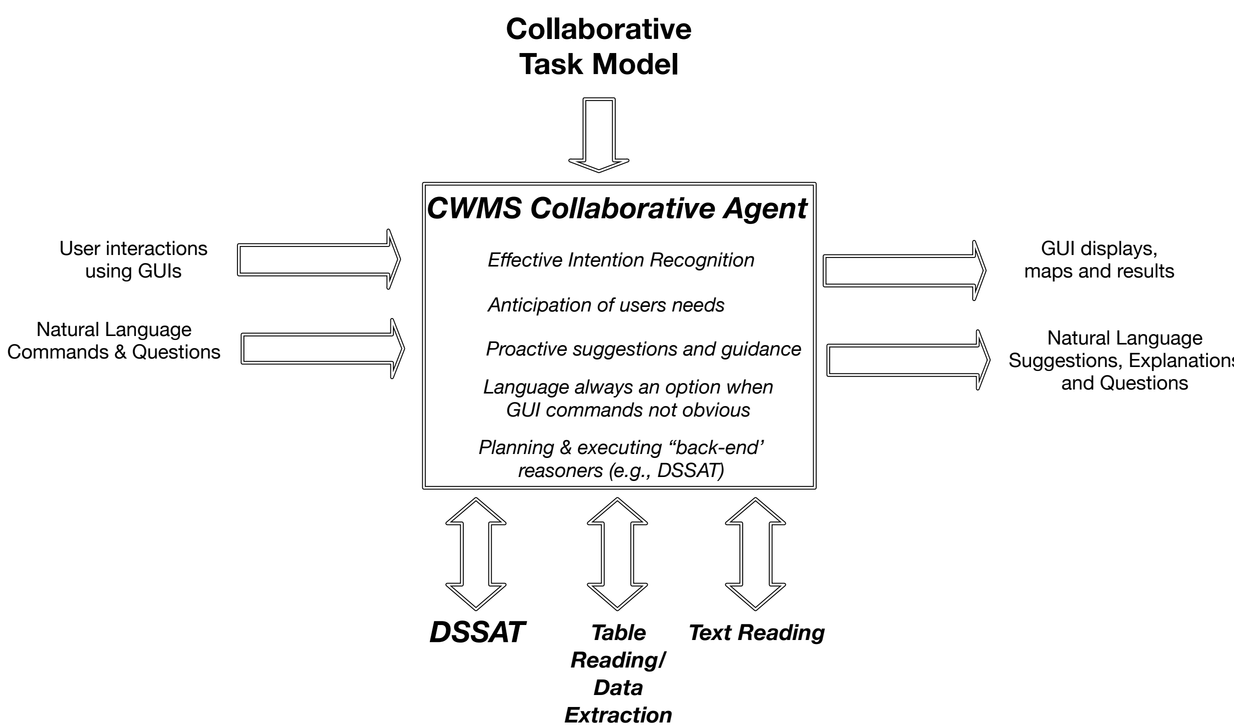

Figure 13. Collaborative Problem Solving Model

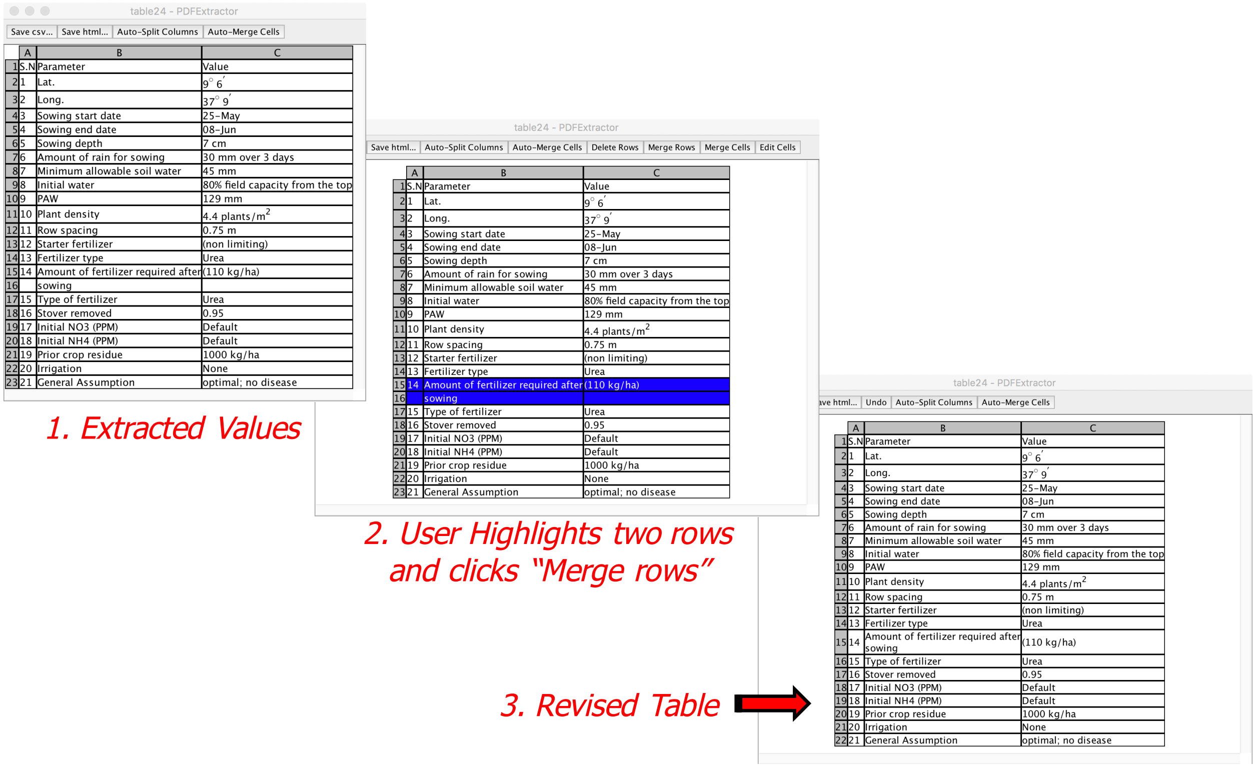

Figure 14. Interactive Table Processing

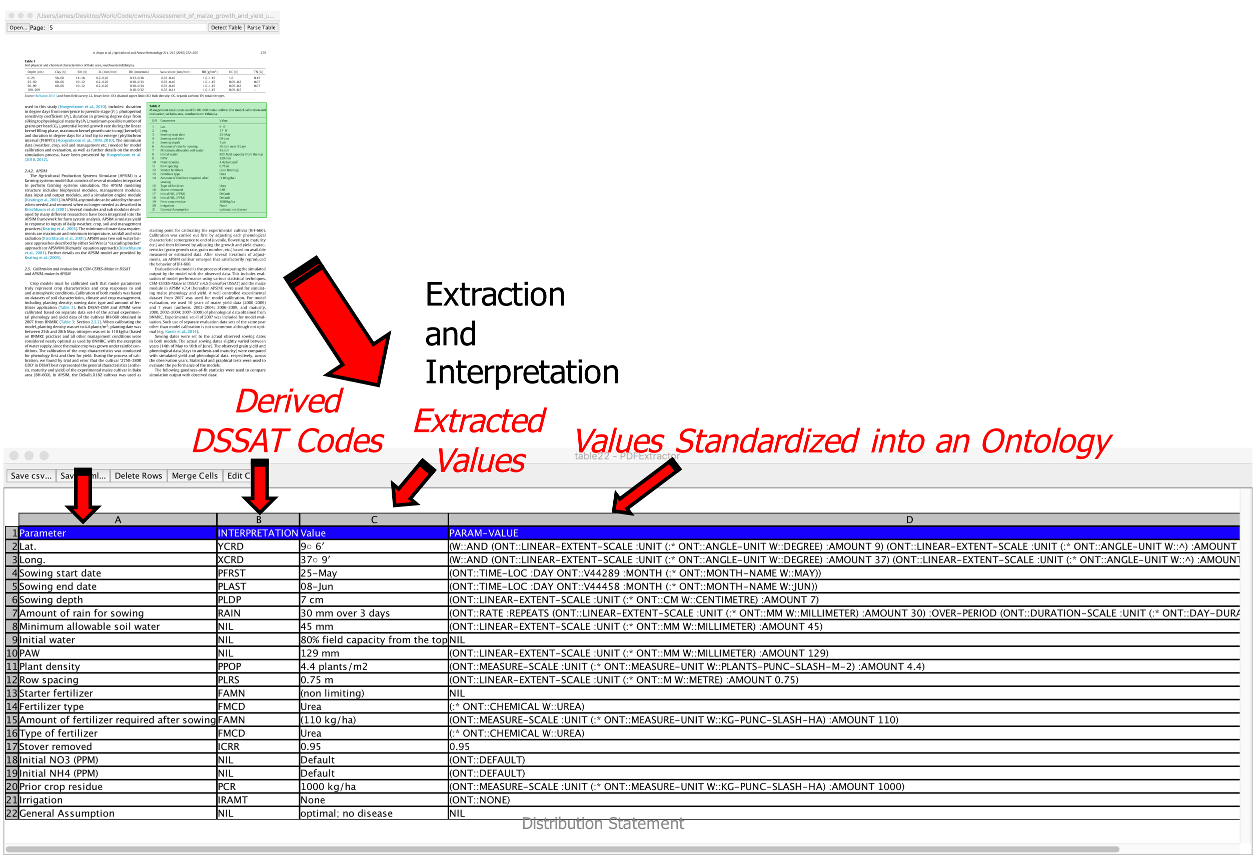

Figure 15. Interpretation of Elements Extracted from a Table

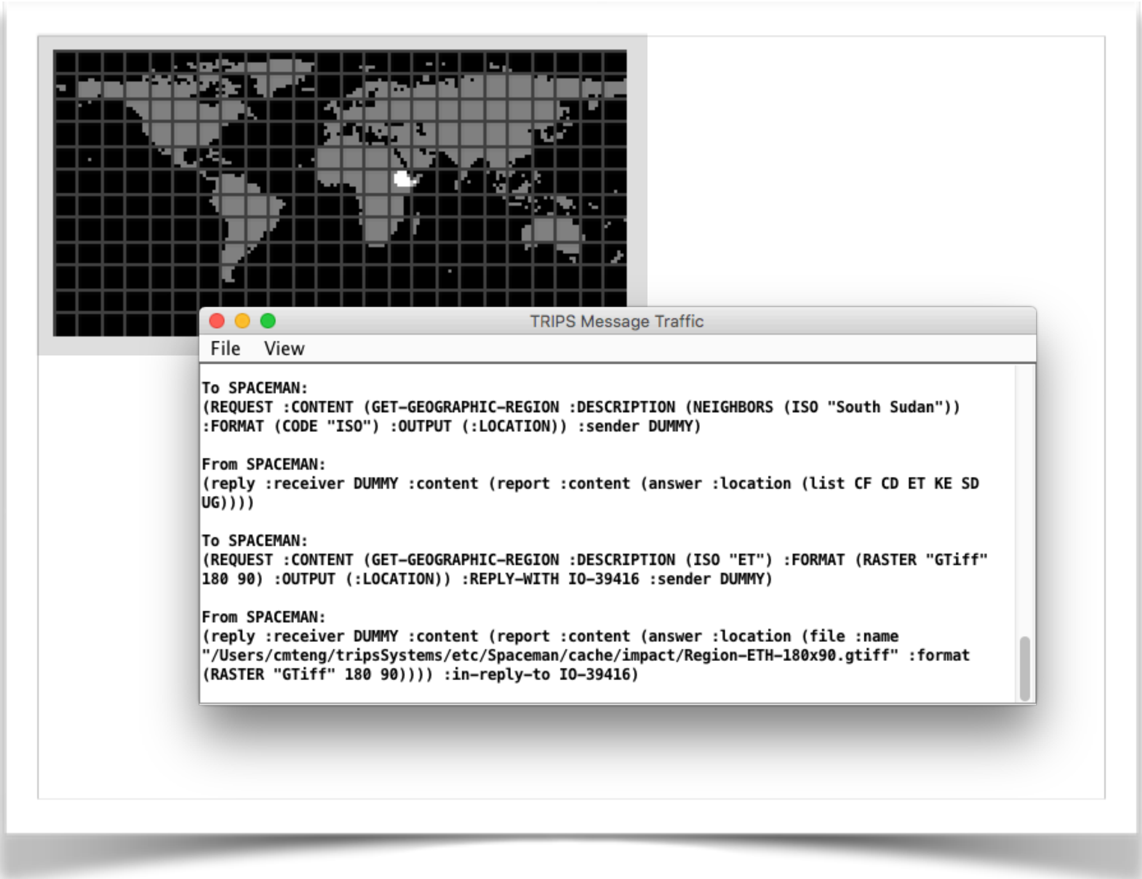

Figure 16. Spaceman Module for handling spatial descriptions in different formats

shows the interpretation of elements extracted from a table. Language fragments are parsed using the TRIPS parser, and the terms are represented in a standardized ontology plus DSSAT codes.

6. SpaceMan

SpaceMan is a module that maps English names and spatial descriptions to locations in a range of different formats used by World Modeling systems. We incorporated a number of external resources for recognizing geonames. The outputs include standardized codes for the locations and visual displays of the locations on a world map. The list of resources integrated into the Spaceman module is given in the Appendix.

We developed a flexible image display component allowing communication using overlays of maps, charts and data. This is critical for effective human understanding of many simulation and experiment results, for instance outputs of DSSAT runs. Figure 16 shows SpaceMan in use in the CWMS system. We packaged and released the code for SpaceMan as a standalone module.

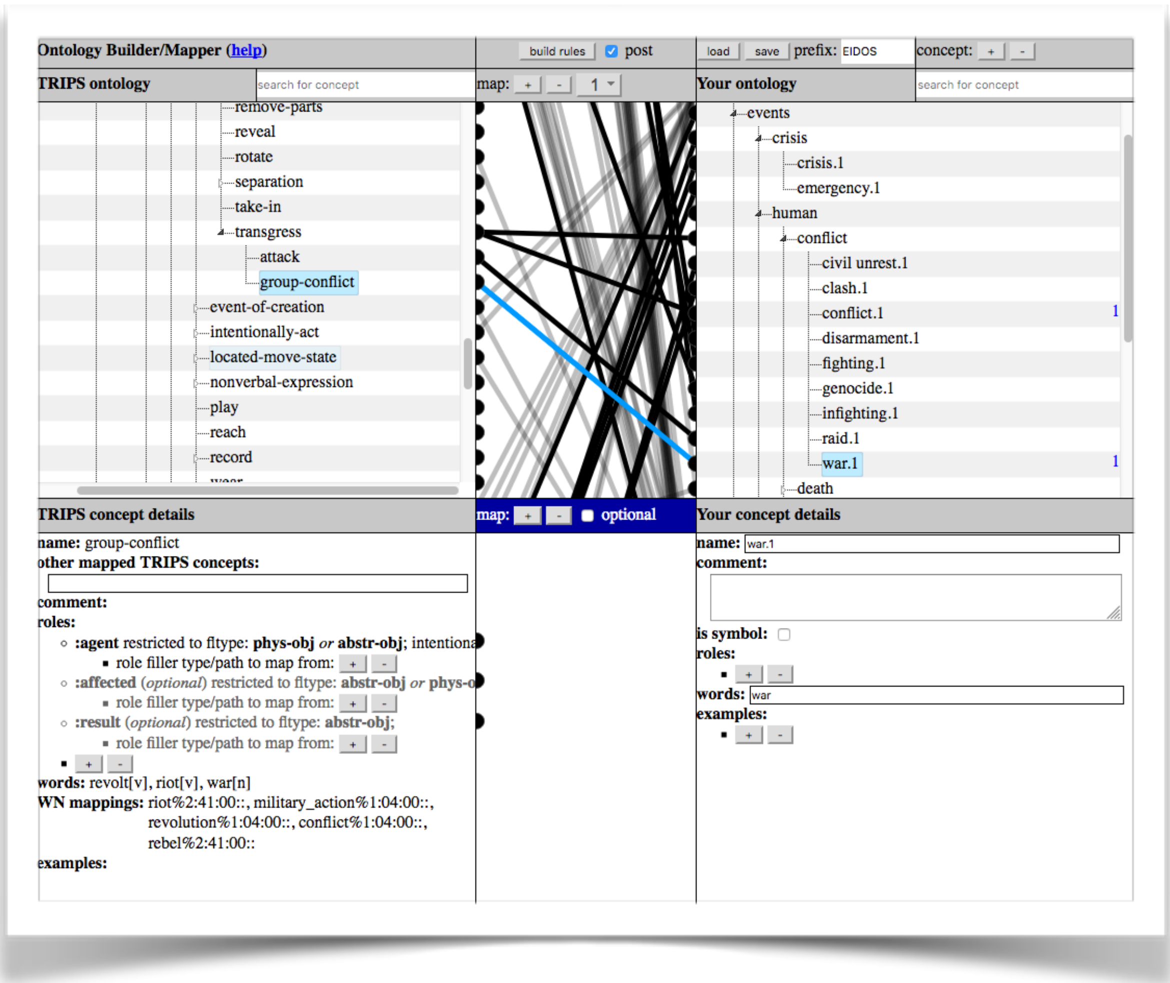

Figure 17. Automatic alignment between the TRIPS and UN ontologies visualized using the ontology mapping tool

7. VariableFinder

Using a spreadsheet of arbitrary ontology specification together with the associated glosses, VariableFinder uses word embedding techniques to map expressions found in documents to a ranked list of types in this ontology. We packaged and released the code for VariableFinder as a standalone module. The VariableFinder method was incorporated into Eidos.

8. Ontology Alignment

We developed a method for automatically aligning different ontologies based on the ontology types and, if available, examples and gloss. The generated alignment can be visualized graphically in a web based tool, and the user can edit the alignment as well as the properties of the ontologies. The amended ontologies and alignment can be saved and reloaded for further analysis and use. Figure 17 shows the alignment between the TRIPS and UN ontologies.

Publications

Allen, J., Teng, C. M. and Galescu, L. (2018). Dialogue as collaborative problem solving: A case study. In *Sixth Annual Conference on Advances in Cognitive Systems*.

Allen, J and Teng, C. M. (2019). Human-machine collaboration for world modeling. In *Modeling the World's Systems*, Washington, DC.

Allen, J., Teng, C. M. and Galescu, L. (2019). Dialogue as collaborative problem solving: A case study. In *Advances in Cognitive Systems*, volume 7, pages 195-214.

Allen, J., Galescu, L., Teng, C. M., and Perera, I. (2020) Conversational agents for complex collaborative tasks. *AI Magazine,* 41(4), pages 54-78.

Resources

- CWMS reader web service: http://trips.ihmc.us/parser/cgi/cwmsreader
- Extended TRIPS ontology and lexicon browser: http://trips.ihmc.us/parser/cgi/lex-ont
- Standalone table reader https://github.com/wdebeaum/PDFExtractor
- SpaceMan: https://github.com/wdebeaum/Spaceman
- VariableFinder: https://github.com/wdebeaum/VariableFinder
- Ontology alignment visualization and editing tool: http://trips.ihmc.us/ontology-mapper/ontology-mapper.html

Appendix A. Resources for recognizing and standardizing geonames

data source name/links	ID prefix	file	maps to TRIPS types	description
mledoze/countries	(none)	`countries.json`	CITY, COUNTRY, GEOGRAPHIC-REGION, NATIONALITY, NATIONALITY-VAL	"This repository contains a list of world countries, as defined by ISO Standard 3166-1" (This is the same data used by the Countries tagger.)
University of Arizona Computational Language Understanding (CLU) Lab supplementary file for GNO derived from the Database of Global Administrative Areas (GADM)	GADM	`gadm_woredas.txt`	DISTRICT	"Eidos has provided a supplementary text file that contains woredas from GADM that are not found in the [GNO] allcountries.zip" "GADM, the Database of Global Administrative Areas, is a high-resolution database of country administrative areas, with a goal of 'all countries, at all levels, at any time period.'"
United States Geological Survey (USGS) Board on Geographic Names (BGN) Geographic Names Information System (GNIS)	GNIS	`NationalFile.txt`	various, mostly under GEOGRAPHIC-REGION and GEO-FORMATION	"The GNIS contains information about physical and cultural geographic features of all types in the United States, associated areas, and Antarctica, current and historical, but not including roads and highways." (This is the same data used by the Terms tagger.)
GeoNames.org	GNO			"The GeoNames geographical database covers all countries and contains over eleven million placenames that are available for download free of charge."
		`admin1CodesASCII.txt`	STATE	"names in English for admin divisions."
		`admin2Codes.txt`	COUNTY	"names for administrative subdivision 'admin2 code'"
		`cities50000.txt`	CITY, DISTRICT	all cities with a population > 50,000 (derived from `cities15000.txt`)
		`ET.txt`, `SS.txt`	various, mostly under GEOGRAPHIC-REGION and GEO-FORMATION	"features for country with iso code XX" (ET=Ethiopia, SS=South Sudan)
National Geospatial-intelligence Agency (NGA) GEONet Names Server (GNS)	GNS	`et.txt`, `od.txt`	various, mostly under GEOGRAPHIC-REGION and GEO-FORMATION	"The GEOnet Names Server (GNS) is the official repository of standard spellings of all foreign geographic names, sanctioned by the United States Board on Geographic Names (US BGN)." (et=Ethiopia, od=South Sudan)
Humanitarian Data Exchange (HDX)	HDX	`eth_pop_adm3.csv`	DISTRICT	"Find, share and use humanitarian data all in one place" (eth=Ethiopia, ssd=South Sudan)
		`eth_populatedplaces_tabulardata.csv`	CITY	
		`ssd_populatedplaces_tabulardata.csv`	CITY	
World Food Programme (WFP) GeoNode	WFP	`eth_bnd_adm2_wfpco.csv`	COUNTY	(eth=Ethiopia, ssd=South Sudan)
		`Ethiopia_bnd_adm2_woreda.csv`	DISTRICT	
		`ssd_ica_mainsettlements_geonode_feb2016.csv`	CITY	
		`ssd_ica_predlhz_geonode_feb2016.csv`	COUNTY	