

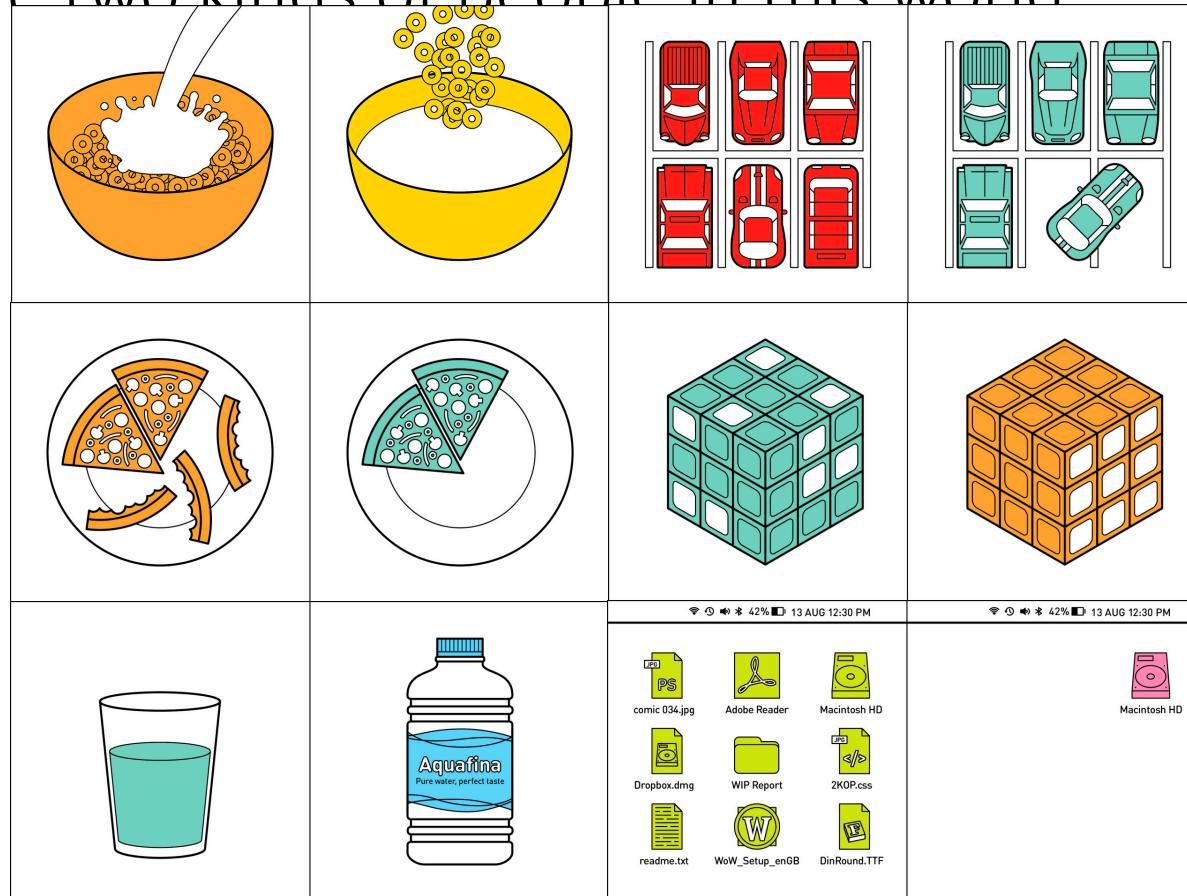
# Structured Data

**Jay Pujara**

**DSCI-558, Spring 2021**

**University of Southern California**

There are two kinds of people in this world



Images courtesy [João Rocha](#)

# There are two kinds of data in this world

## Structured Data

Project	Data Exposed	Size of Dump and Data Set	Archive URL
Addgene	Addgene catalog (tab delimited file)	1.1 MB	<a href="#">tab-delimited file</a>
Allen Brain Atlas	Science Commons extract from ABA Web site, on or shortly before 26 Feb 2007	51 MB	<a href="#">dump file</a>
Airport Data	SPARQL	754,585	
BAMS	BAMS	5.6 MB	<a href="#">bams-from-swanson-98-4-23-07.owl</a>
BBC John Peel sessions	holding data released during Hackday, 2007	277,000 triples	<a href="#">peel.tar.gz</a>
BBOP	All OBO ontologies	36 MB	<a href="#">obo-all.tgz</a>
BBOP	selected OBO ontologies, downloaded ~21 April 2007, augmented with inferred relations	2.6 MB	<a href="#">obo-in-owl.tgz</a>
Billion Triples Challenge Dataset 2008	various dumps	1 billion triples	<a href="#">download page</a>
Billion Triples Challenge Dataset 2009	crawled Web data	1.14 billion triples	<a href="#">download page</a>
Billion Triples Challenge Dataset 2010	crawled Web data	3.2 billion triples	<a href="#">download page</a>

## Unstructured Data

### 1.2.1 Structure Enables Sophisticated Processing

A key factor in the re-usability of data is the extent to which it is well *structured*. The more regular and v

While most Web sites have some degree of structure, the language in which they are created, HTML, is for software applications to extract snippets of structured data from HTML pages.

To address this issue, a variety of *microformats* have been invented. Microformats can be used to pub through embedding of data in HTML pages. As microformats tightly specify how to embed data, applica representing data about a small set of different types of entities; they only provide a small set of attribut for example, that a person is the speaker of an event, rather than being just an attendee or the organiz

A more generic approach to making structured data available on the Web are *Web APIs*. Web APIs pro *Product Advertising API* and the *Flickr API*. The site *ProgrammableWeb* maintains a directory conta

The advent of Web APIs has led to an explosion in small, specialized applications (or *mashups*) that co of programmatic access to structured data are indisputable, the existence of a specialized API for each programmer must understand the methods available to retrieve data from each API, and write custom c

### 1.2.2 Hyperlinks Connect Distributed Data

It is common for Web APIs to provide results in structured data formats such as XML and JSON, whic limitations, which are best explained by comparison with HTML. The HTML specification defines the *an*user agents, such as browsers and search engine cr can follow, or to traverse the link directly in order to retrieve and process the referenced document. It is documents. By contrast, the data returned from the majority of Web APIs does not have the equivalent

Furthermore, many Web APIs refer to items of interest using identifiers that have only local scope – e.g standard mechanism to refer to items described by one API in data returned by another.

Consequently, data returned from Web APIs typically exists as isolated fragments, lacking reliable onw; it truly *in the Web*, making it linkable and therefore discoverable.

To return to the comparison with HTML, the analogous situation would be a search engine that require publisher would need to register each Web page with each search engine. The ability for anyone to ad with browsers, have historically been key drivers of the Web's explosive growth. The same principles of realize such linkage.

# Sources of Structured Data

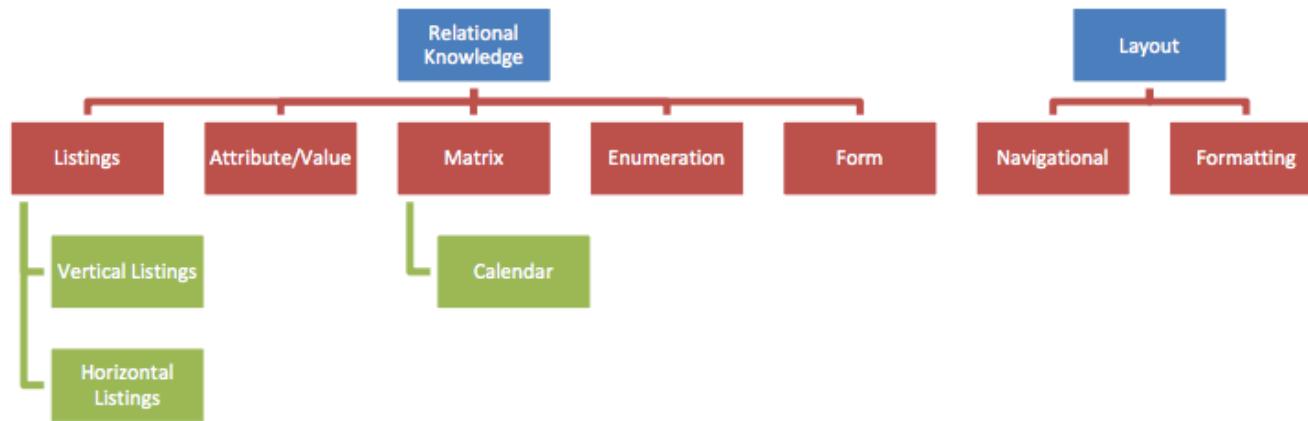
- Knowledge Graphs!
- The Semantic Web
- Databases
- Structured files (CSV, XML, XLS)
- Structures in HTML, e.g. tables

# Structured Data: Overview of Topics

- Structured Tables on the Web
- Extracting Relational Data
- Extracting Quantitative Data

# **Structured Data** In Tables

# Census of tables on the Web



From Web-Scale Table Census and Classification (Crestan & Pantel, WSDM11)

# Horizontal Listings

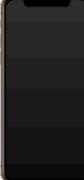
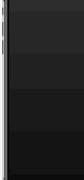
iPhone	Released with	Release date	Final supported OS	Support ended
<a href="#">iPhone</a>	<a href="#">iPhone OS 1.0</a>	June 29, 2007	<a href="#">iPhone OS 3.1.3</a>	June 20, 2010
<a href="#">iPhone 3G</a>	<a href="#">iPhone OS 2.0</a>	July 11, 2008	<a href="#">iOS 4.2.1</a>	March 3, 2011
<a href="#">iPhone 3GS</a>	<a href="#">iPhone OS 3.0</a>	June 19, 2009	<a href="#">iOS 6.1.6</a>	September 18, 2013
<a href="#">iPhone 4</a>	<a href="#">iOS 4.0</a>	June 21, 2010	<a href="#">iOS 7.1.2</a>	September 17, 2014
<a href="#">iPhone 4S</a>	<a href="#">iOS 5.0</a>	October 14, 2011	<a href="#">iOS 9.3.5</a>	September 12, 2016
<a href="#">iPhone 5</a>	<a href="#">iOS 6.0</a>	September 21, 2012	<a href="#">iOS 10.3.3</a>	September 18, 2017
<a href="#">iPhone 5C</a>	<a href="#">iOS 7.0</a>	September 20, 2013	<a href="#">iOS 10.3.3</a>	September 18, 2017
<a href="#">iPhone 5S</a>	<a href="#">iOS 7.0</a>	September 20, 2013	<a href="#">latest iOS</a>	(current)
<a href="#">iPhone 6 (Plus)</a>	<a href="#">iOS 8.0</a>	September 19, 2014	<a href="#">latest iOS</a>	(current)
<a href="#">iPhone 6S (Plus)</a>	<a href="#">iOS 9.0</a>	September 25, 2015	<a href="#">latest iOS</a>	(current)
<a href="#">iPhone SE</a>	<a href="#">iOS 9.3</a>	March 31, 2016	<a href="#">latest iOS</a>	(current)
<a href="#">iPhone 7 (Plus)</a>	<a href="#">iOS 10.0</a>	September 16, 2016	<a href="#">latest iOS</a>	(current)
<a href="#">iPhone 8 (Plus)</a>	<a href="#">iOS 11.0</a>	September 22, 2017	<a href="#">latest iOS</a>	(current)
<a href="#">iPhone X</a>	<a href="#">iOS 11.0.1</a>	November 3, 2017	<a href="#">latest iOS</a>	(current)
<a href="#">iPhone XS (Max)</a>	<a href="#">iOS 12</a>	September 21, 2018	<a href="#">latest iOS</a>	(current)
<a href="#">iPhone XR</a>	<a href="#">iOS 12</a>	October 26, 2018	<a href="#">latest iOS</a>	(current)

# Horizontal Listings

## Elections in the 1880s [edit]

Edinburgh West by-election, 1888 [32][24]					
Party	Candidate	Votes	%	±	
Liberal	Thomas Buchanan	3,294	50.4	+6.7	
Liberal Unionist	Thomas Raleigh	3,248	49.6	-6.7	
		Majority	46	0.8	N/A
		Turnout	6,542	84.4	+12.0
		Registered electors	7,749		
Liberal	gain from Liberal Unionist	Swing	+6.7		

# Vertical Listings

iPhone 8	iPhone 8 Plus	iPhone XS	iPhone XS Max	iPhone XR
				
IOS 11.0			IOS 12.0	
		IOS 12.0.1		
		IOS 12.1		
In addition to 7: True Tone display	In addition to 7 Plus: True Tone display	5.85 in (149 mm), 5.31 in (135 mm) by 2.45 in (62 mm), ~19.5:9 aspect ratio, aluminosilicate glass covered 16,777,216-color (24-bit), AMOLED screen, 2,436 x 1,125 px screen resolution at 458 ppi, 1,000,000:1 contrast ratio, 625 cd/m² max brightness, fingerprint-resistant oleophobic coating, True Tone display, Dolby Vision and HDR10 support	6.46 in (164 mm), 5.9 in (150 mm) by 2.73 in (69 mm), ~19.5:9 aspect ratio, aluminosilicate glass covered 16,777,216-color (24-bit), AMOLED screen, 2,688 x 1,242 px screen resolution at 458 ppi, 1,000,000:1 contrast ratio, 625 cd/m² max brightness, fingerprint-resistant oleophobic coating, True Tone display, Dolby Vision and HDR10 support	6.06 in (154 mm), 5.54 in (141 mm) by 2.56 in (65 mm), ~19.5:9 aspect ratio, aluminosilicate glass covered 16,777,216-color (24-bit), LED backlight screen, 1,792 x 828 px screen resolution at 326 ppi, 1,400:1 contrast ratio, 625 cd/m² max brightness, fingerprint-resistant oleophobic coating, True Tone display, Dolby Vision and HDR10 support
64 and 256 GB NAND Flash driven by NVMe Express controller		64, 256, and 512 GB NAND Flash driven by NVMe Express controller		64, 128, and 256 GB NAND Flash driven by NVMe Express controller
2.39 GHz hexa-core Apple-designed 64-bit Apple A11 Bionic (6-cores: 2 Monsoon high-performance, 4 Mistral high-efficiency) with embedded M11 motion coprocessor and dual-core Neural Engine		2.49 GHz hexa-core Apple-designed 64-bit Apple A12 Bionic (6-cores: 2 Vortex high-performance, 4 Tempest high-efficiency) with octa-core Neural Engine		
64-bit <sup>[8]</sup>				
Apple designed tri-core GPU		Apple designed (1.1GHz, quad-core) "G11P" GPU <sup>[8]</sup>		
2 GB LPDDR4X DRAM	3 GB LPDDR4X DRAM	4 GB LPDDR4X DRAM	3 GB LPDDR4X DRAM	
8-pin Lightning connector				
Wi-Fi (802.11 a/b/g/n/ac) with two spatial stream MIMO				
Nano-SIM and eSIM <sup>[10]</sup>				
x-SIM	Nano-SIM in China <sup>[note 1][11][12][13]</sup>	Dual Nano-SIM in China, Hong Kong and Macau <sup>[11][12][13]</sup>		

# Attribute/Value Tables

<b>Brand</b>	Apple Inc.
<b>Manufacturer</b>	Foxconn <sup>[1]</sup> (on contract) Pegatron <sup>[1]</sup> (on contract)
<b>Slogan</b>	<i>Say hello to the future</i>
<b>Generation</b>	11th
<b>Model</b>	A1865 (with Qualcomm modem) A1901 (with Intel modem) A1902 (sold in Japan) <sup>[2]</sup>
<b>Compatible networks</b>	GSM, CDMA2000, EV-DO, HSPA+, LTE, LTE Advanced
<b>First released</b>	November 3, 2017
<b>Availability by Region</b>	November 3, 2017 [show] November 23, 2017 [show] November 24, 2017 [show] December 1, 2017 [show] December 7, 2017 [show] December 8, 2017 [show] December 22, 2017 [show]
<b>Discontinued</b>	September 12, 2018
<b>Predecessor</b>	<a href="#">iPhone 7 / iPhone 7 Plus</a>
<b>Successor</b>	<a href="#">iPhone XS / iPhone XS Max</a> <a href="#">iPhone XR</a>
<b>Type</b>	Smartphone
<b>Form factor</b>	Slate
<b>Dimensions</b>	H: 143.6 mm (5.65 in) W: 70.9 mm (2.79 in) D: 7.7 mm (0.30 in)

# Matrix

**U.S. Exports of Crude Oil (Thousand Barrels)**

<b>Year</b>	<b>Jan</b>	<b>Feb</b>	<b>Mar</b>	<b>Apr</b>	<b>May</b>	<b>Jun</b>	<b>Jul</b>	<b>Aug</b>	<b>Sep</b>	<b>Oct</b>	<b>Nov</b>	<b>Dec</b>
<b>1920</b>	469	853	892	693	761	627	723	553	790	777	796	823
<b>1921</b>	743	794	750	748	874	586	538	885	881	747	869	525
<b>1922</b>	727	583	806	924	771	826	893	764	1,127	741	879	1,122
<b>1923</b>	762	666	1,028	1,511	1,324	2,598	1,547	1,542	1,592	1,315	1,397	2,103
<b>1924</b>	1,528	1,680	1,550	1,547	1,858	1,542	1,409	1,242	1,893	1,488	1,453	1,049
<b>1925</b>	1,149	1,122	1,058	798	1,356	1,255	1,302	1,465	923	1,292	740	877
<b>1926</b>	1,183	1,049	965	1,308	1,842	1,226	1,726	1,083	1,388	1,010	1,344	1,283
<b>1927</b>	1,204	1,165	1,199	1,171	1,390	1,411	1,089	1,382	1,297	1,539	1,280	1,717
<b>1928</b>	1,225	1,243	1,530	1,303	1,493	1,879	1,669	1,883	1,506	2,015	1,691	1,529
<b>1929</b>	1,972	1,678	1,600	1,726	1,932	2,615	3,117	2,236	1,988	2,869	2,579	2,089
<b>1930</b>	1,808	1,731	1,944	1,900	2,202	2,508	1,973	2,407	1,961	2,167	1,765	1,339
<b>1931</b>	1,919	1,710	1,586	1,826	2,268	2,544	2,621	2,856	2,296	2,389	2,449	1,071
<b>1932</b>	1,592	1,897	2,090	2,867	2,942	2,791	2,249	2,839	2,113	2,541	1,318	2,154
<b>1933</b>	1,913	1,886	2,137	2,939	2,679	4,355	4,523	3,141	3,182	3,888	3,305	2,636
<b>1934</b>	2,288	2,511	2,582	3,942	3,724	3,794	4,128	3,696	4,068	3,277	4,680	2,437
<b>1935</b>	2,369	2,804	3,281	3,776	4,613	5,589	5,832	4,946	4,971	4,810	4,289	4,098
<b>1936</b>	3,067	3,474	3,155	3,743	4,390	4,792	4,458	5,561	5,025	4,708	4,145	3,666
<b>1937</b>	3,596	3,777	3,196	4,899	6,796	6,181	6,363	7,423	6,602	6,692	6,645	5,116
<b>1938</b>	5,953	5,328	6,121	7,553	7,798	7,424	7,250	7,003	5,577	6,780	5,602	4,884
<b>1939</b>	4.477	4.810	4.966	6.222	8.643	5.831	7.304	5.969	6.925	6.947	5.323	4.656

# Matrix - Calendar

SUNDAY	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY
30  Mostly Sunny Actual: 76°   61° 0 in	1  Partly Cloudy Actual: 83°   66° 0 in	2  Partly Sunny Actual: 76°   67° 0 in	3  Rain Actual: 77°   67° 0 in	4  Partly Cloudy Actual: 75°   65° 0 in	5  Partly Cloudy Actual: 73°   65° 0 in	6  Partly Sunny Actual: 75°   63° 0 in
7  Partly Cloudy Actual: 74°   61° 0 in	8  Partly Cloudy Actual: 73°   64° 0 in	9  Partly Cloudy Actual: 74°   64° 0 in	10  Partly Cloudy Actual: 72°   62° 0 in	11  Mostly Sunny Forecast: 74°   61° 0 in	12  Sunny Forecast: 76°   64° 0.01 in	13  Mostly Cloudy Forecast: 68°   61° 0 in
14  Partly Cloudy Forecast: 68°   60° 0 in	15  Sunny Forecast: 77°   57° 0 in	16  Sunny Forecast: 76°   59° 0 in	17  Sunny Forecast: 78°   59° 0 in	18  Sunny Forecast: 78°   60° 0 in	19  Sunny Forecast: 78°   61° 0 in	20  Sunny Forecast: 78°   62° 0 in
21  Sunny Forecast: 77°   63° 0 in	22  Sunny Forecast: 76°   64° 0 in	23  Sunny Forecast: 75°   63° 0 in	24  Sunny Forecast: 74°   63° 0 in	25 Average: 72°   58° 0 in	26 Average: 72°   58° 0 in	27 Average: 72°   57° 0 in
28 Average: 72°   57° 0 in	29 Average: 72°   57° 0 in	30 Average: 72°   57° 0 in	31 Average: 72°   57° 0 in	1 Average: 71°   56° 0 in	2 Average: 71°   56° 0 in	3

# Enumeration

Classes in the OWL vocabulary:

<b>rdfs:Class</b>
<a href="#">owl:AllDifferent</a>
<a href="#">owl:AnnotationProperty</a>
<a href="#">owl:Class</a>
<a href="#">owl:DataRange</a>
<a href="#">owl:DatatypeProperty</a>
<a href="#">owl:DeprecatedClass</a>
<a href="#">owl:DeprecatedProperty</a>
<a href="#">owl:FunctionalProperty</a>
<a href="#">owl:InverseFunctionalProperty</a>
<a href="#">owl:Nothing</a>
<a href="#">owl:ObjectProperty</a>
<a href="#">owl:Ontology</a>
<a href="#">owl:OntologyProperty</a>
<a href="#">owl:Restriction</a>
<a href="#">owl:SymmetricProperty</a>
<a href="#">owl:Thing</a>
<a href="#">owl:TransitiveProperty</a>

# Other

	First Round		Quarterfinals		Semifinals		Final	
1	D Brown P Marx	6 6	1	D Brown P Marx	4 3	Alt	M Crugnola S Ianni	6 6
	T Gabashvili B Pašanski	4 0				Alt	M Crugnola S Ianni	6 6
Alt	M Crugnola S Ianni	5 6 [10]				Alt	M Crugnola S Ianni	6 6
	A Motti P Raja	7 4 [4]				4	M Draganja M Pavić	4 4
4	M Draganja M Pavić	6 6	4	M Draganja M Pavić	7 3 [18]		M Crugnola S Ianni	5 6 <sup>2</sup>
WC	Davide Borromini Leandro Mosconi	1 0		H Kontinen A Siljeström	6 <sup>5</sup> 6 [16]		R Junaid I Zelenay	7 7 <sup>7</sup>
WC	Carlo Alberto Fossati Christian Scandroglio	0 0						
	H Kontinen A Siljeström	6 6						
WC	D Giorgini Matteo Volante	6 <sup>4</sup> 3						
	A Satschko A Shamasdin	7 <sup>7</sup> 6						
	I Bozoljac D Lajović	6 3 [8]	3	A Satschko A Shamasdin	7 6		A Satschko A Shamasdin	1 3
3	F Moser F Škugor	3 6 [10]		F Moser F Škugor	5 4		R Junaid I Zelenay	6 6
	R Olivo M Trungelliti	4 4						
	A Giannessi P Starace	6 6						
	J Aguilar E López-Pérez	3 2	2	A Giannessi P Starace	5 3			
2	R Junaid I Zelenay	6 6		R Junaid I Zelenay	7 6			

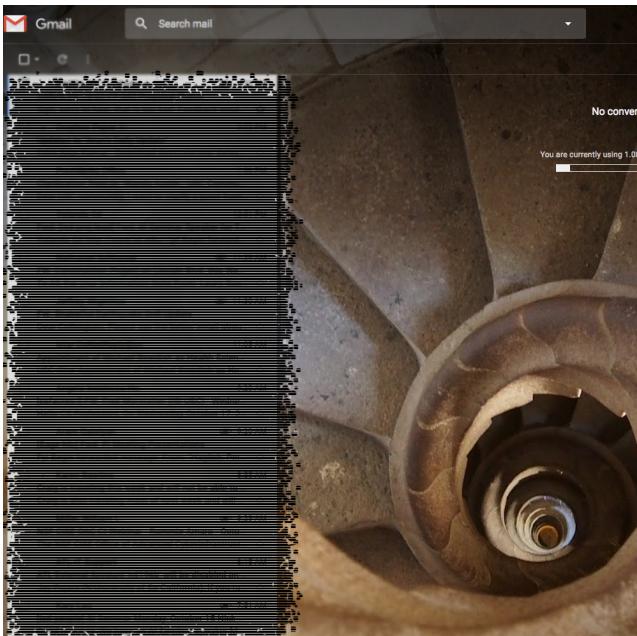
# Navigational

## TABLE OF CONTENTS

- [1. Features](#)
- [2. Linguistic annotations](#)
- [3. Tokenization](#)
- [4. POS tags and dependencies](#)
- [5. Named entities](#)
- [6. Word vectors and similarity](#)
- [7. Pipelines](#)
- [8. Vocab, hashes and lexemes](#)

<b>Amazon Music</b> Stream millions of songs	<b>Amazon Drive</b> Cloud storage from Amazon	<b>6pm</b> Score deals on fashion brands	<b>AbeBooks</b> Books, art & collectibles	<b>ACX</b> Audiobook Publishing Made Easy	<b>Alexa</b> Actionable Analytics for the Web	<b>Amazon Business</b> Everything For Your Business
<b>AmazonFresh</b> Groceries & More Right To Your Door	<b>AmazonGlobal</b> Ship Orders Internationally	<b>Home Services</b> Handpicked Pros Happiness Guarantee	<b>Amazon Inspire</b> Digital Educational Resources	<b>Amazon Rapids</b> Fun stories for kids on the go	<b>Amazon Restaurants</b> Food delivery from local restaurants	<b>Amazon Web Services</b> Scalable Cloud Computing Services
<b>Audible</b> Download Audiobooks	<b>Book Depository</b> Books With Free Delivery Worldwide	<b>Box Office Mojo</b> Find Movie Box Office Data	<b>ComiXology</b> Thousands of Digital Comics	<b>CreateSpace</b> Indie Print Publishing Made Easy	<b>DPRReview</b> Digital Photography	<b>East Dane</b> Designer Men's Fashion
<b>Fabric</b> Sewing, Quilting & Knitting	<b>Goodreads</b> Book reviews & recommendations	<b>IMDb</b> Movies, TV & Celebrities	<b>IMDbPro</b> Get Info Entertainment Professionals Need	<b>Junglee.com</b> Shop Online in India	<b>Kindle Direct Publishing</b> Indie Digital Publishing Made Easy	<b>Prime Now</b> Ultrafast Delivery on Everyday Items
<b>Amazon Photos</b> Unlimited Photo Storage Free With Prime	<b>Prime Video Direct</b> Video Distribution Made Easy	<b>Shopbop</b> Designer Fashion Brands	<b>TenMarks.com</b> Math Activities for Kids & Schools	<b>Amazon Warehouse</b> Great Deals on Quality Used Products	<b>Whole Foods Market</b> America's Healthiest Grocery Store	<b>Withoutabox</b> Submit to Film Festivals
	<b>Woot!</b> Deals and Shenanigans	<b>Zappos</b> Shoes & Clothing	<b>Souq.com</b> Shop Online in the Middle East	<b>Subscribe with Amazon</b> Discover & try subscription services		

# Formatting



USC Viterbi

Department of  
Computer Science

USC University of  
Southern California

Contact Google Custom Search



About Research People Search Academic Programs Student Resources Admission



Cue the Robots USC sends computer science attendees to the Grace Hopper Conference, the USC Robotics Open House leads the way for the next generation of STEM professionals.

[Read More](#)

## Quick Links

<a href="#">CS@CS Institutes</a>	<a href="#">Industry Affiliate Program</a>	<a href="#">Giving</a>	<a href="#">Apply Now</a>
<a href="#">SAL Conference Rooms</a>	<a href="#">myViterbi (USC Portal)</a>	<a href="#">Student Resources</a>	<a href="#">Open Faculty Positions</a>

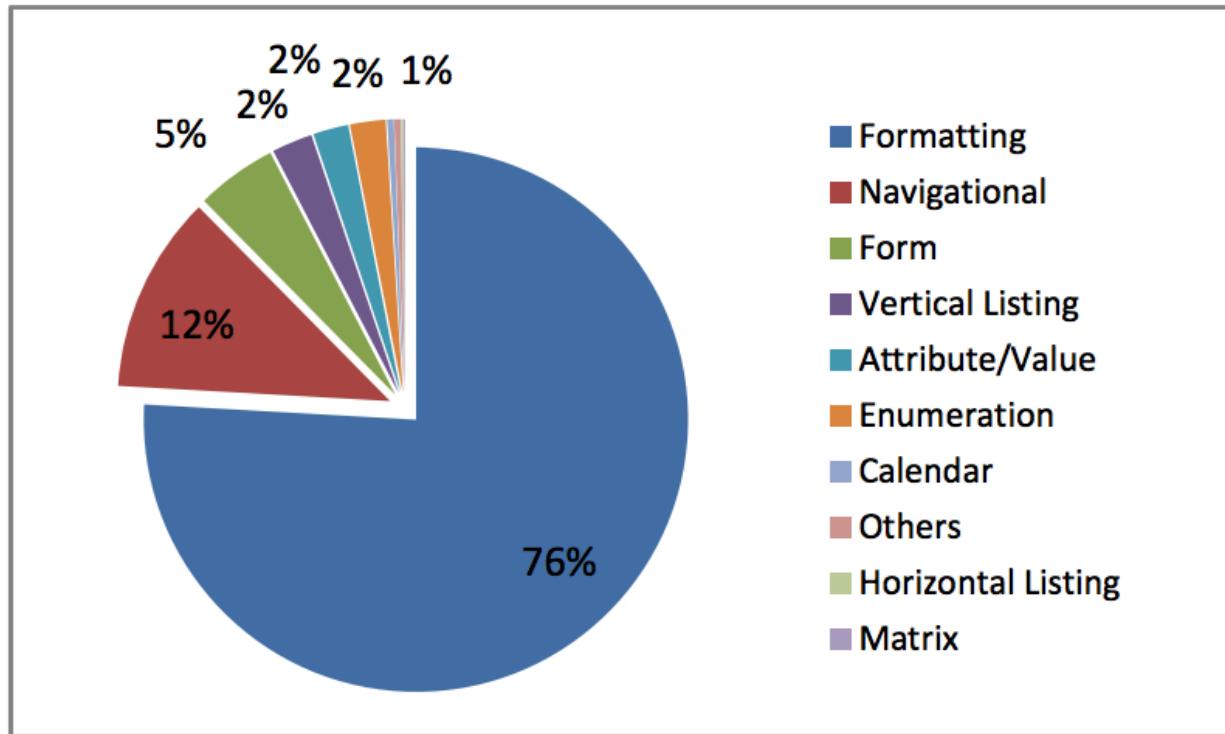
## Events

OCT 15 CS Colloquium: Judea Pearl (UCLA) - The New Science of Cause and Effect  
MON Mon, Oct 15, 2018  
12:00 PM - 1:00 PM  
Location: Wallis Annenberg Hall (ANN) 106 (Sheindlin Forum)  
Speaker: Professor Judea Pearl, UCLA Talk Title: The New Science of Cause and Effect Series: Computer Science Colloquium Abstract:  
Professor Judea Pearl's talk will summarize a revolution that has changed the way scientists deal with...

## News

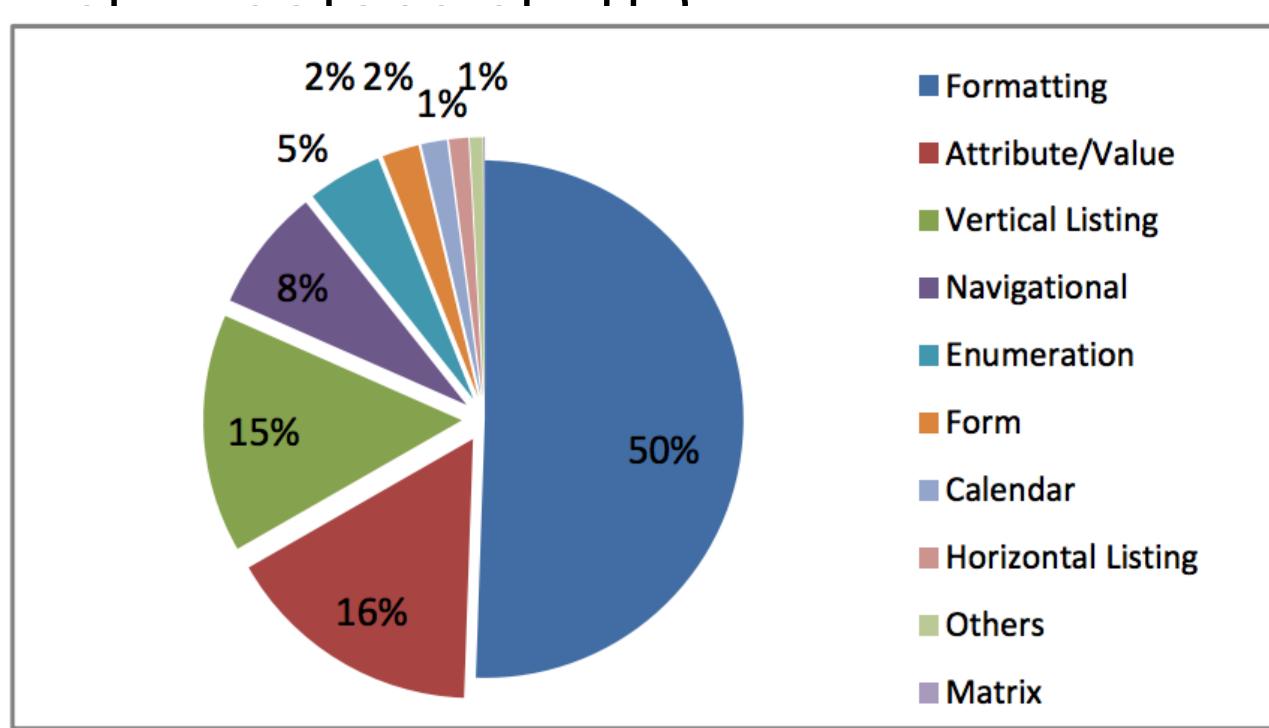


# What's out there?



From Web-Scale Table Census and Classification (Crestan & Pantel, WSDM11)

What's out there after we filter out the obvious



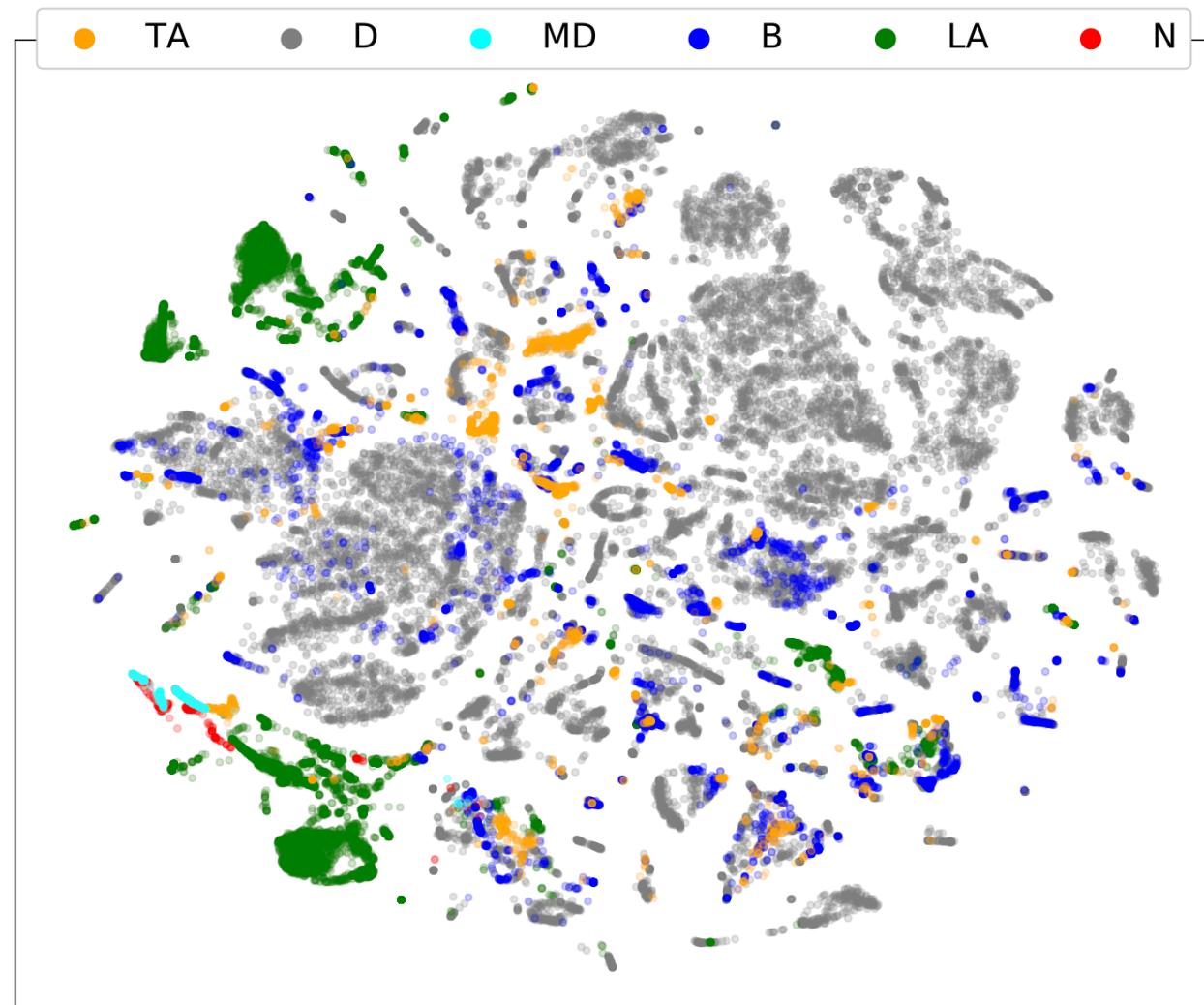
From Web-Scale Table Census and Classification (Crestan & Pantel, WSDM11)

# Automatically determining table types

- Global Layout features
  - #rows, #cols, max cell length
- Layout features (per row/column)
  - cell length, length variance, ratio in col/rowspan
- HTML features (content)
  - unique tags, ratio of th/anchor/img/input;br/font...
- Cell features (content)
  - unique strings, end-with-colon, is/contains number, non-blank

Reference	System											Total
	Formatting	Navigational	Attribute/Value	Vertical Listing	Horizontal Listing	Enumeration	Calendar	Matrix	Form	Others		
Formatting	2235	55	101	67	3	44	0	1	23	3	2532	
Navigational	183	135	5	13	2	41	1	0	1	2	383	
Attribute/Value	85	5	622	84	7	4	0	1	1	2	811	
Vertical Listing	79	10	69	535	14	20	2	3	1	10	743	
Horizontal Listing	14	2	14	17	8	0	1	1	0	1	58	
Enumeration	70	50	9	26	2	75	0	0	1	0	233	
Calendar	1	0	0	7	0	0	69	1	0	0	78	
Matrix	0	0	1	3	0	0	1	2	0	0	7	
Form	38	2	4	1	0	0	0	0	0	1	113	
Others	13	0	4	13	0	0	0	0	0	0	12	42
Total	2718	259	829	766	36	184	74	9	94	31	5000	

From Web-Scale Table Census and Classification  
(Crestan & Pantel,  
WSDM11)

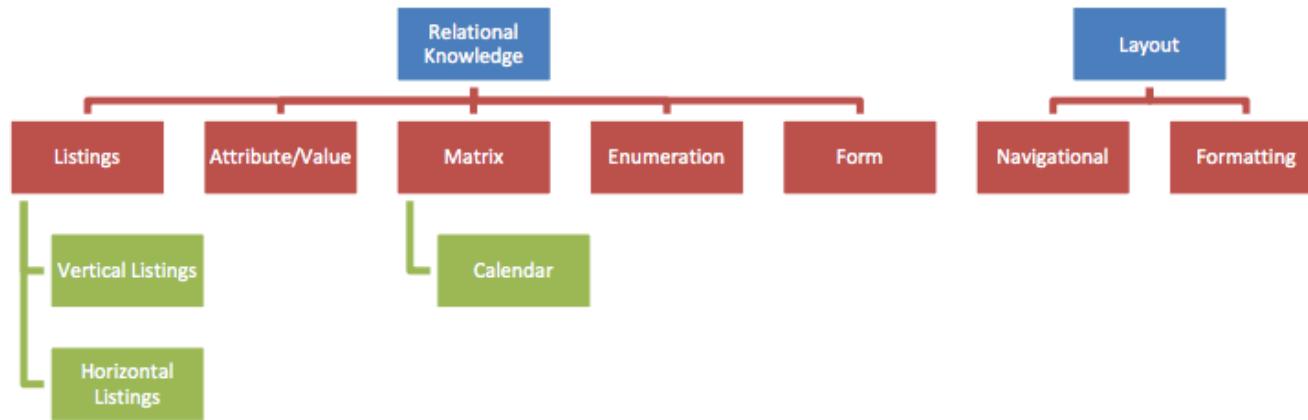


# Extracting Relational Data from Tables

# Modeling Relational Data in Tables

- Find the structural relationships in each table type
- Extract RDF triples from tabular data
  - What is the semantic type of each column?
    - Semantic Labeling
  - What is the subject of the table?
    - Entity Linking
  - What predicates describe the relationships in the table?
    - Semantic Modeling

# Census of tables on the Web



From Web-Scale Table Census and Classification (Crestan & Pantel, WSDM11)

# Horizontal Listings

Predicate

iPhone	Released with	Release date	Final supported OS	Support ended
iPhone	iPhone OS 1.0	June 29, 2007	iPhone OS 3.1.3	June 20, 2010
iPhone 3G	iPhone OS 2.0	July 11, 2008	iOS 4.2.1	March 3, 2011
iPhone 3GS	iPhone OS 3.0	June 19, 2009	iOS 6.1.6	September 18, 2013
iPhone 4	iOS 4.0	June 21, 2010	iOS 7.1.2	September 17, 2014
iPhone 4S	iOS 5.0	October 14, 2011	iOS 9.3.5	September 12, 2016
iPhone 5	iOS 6.0	Rows are records		
iPhone 5C	iOS 7.0			
iPhone 5S	iOS 7.0	September 20, 2013	latest iOS	(current)
iPhone 6 (Plus)	iOS 8.0	September 19, 2014	latest iOS	(current)
iPhone 6S (Plus)	iOS 9.0	September 25, 2015	latest iOS	(current)
iPhone SE	iOS 9.3	March 31, 2016	latest iOS	(current)
iPhone 7 (Plus)	iOS 10.0	September 16, 2016	latest iOS	(current)
iPhone 8 (Plus)	iOS 11.0	September 22, 2017	latest iOS	(current)
iPhone X	iOS 11.0.1	November 3, 2017	latest iOS	(current)
iPhone XS (Max)	iOS 12	September 21, 2018	latest iOS	(current)
iPhone XR	iOS 12	October 26, 2018	latest iOS	(current)

Subject

Object

# Horizontal Listings

Elections in the 1880s <a href="#">[edit]</a>		Subject			
Predicate		Candidate	Votes	%	±
	Party				
	Liberal	Thomas Buchanan	3,294	50.4	+6.7
	Liberal Unionist	Thomas Raleigh	2,248	39.6	-6.7
		Majority	46	0.8	N/A
		Turnout	6,542	84.4	+12.0
		Registered electors	7,749		
		Swing	+6.7		
	Liberal gain from Liberal Unionist				

Object

Other junk?

# Horizontal Listings can be complicated

Subject

Elections in the 1880s [edit]

Edinburgh West by-election, 1888 [32][24]					
Party	Candidate	Votes	%	±	
Liberal	Thomas Buchanan	3,294	50.4	+6.7	
Liberal Unionist	Thomas Raleigh	3,248	49.6	-6.7	
	Majority	46	0.8	N/A	
	Turnout	6,542	84.4	+12.0	
	Registered electors	7,749			
	Swing		+6.7		
Liberal gain from Liberal Unionist					

Predicate

Object

The diagram illustrates the structure of a horizontal list. The table is divided into three main sections: Subject (top), Predicate (middle left), and Object (middle right). The Subject section highlights the title "Edinburgh West by-election, 1888". The Predicate section highlights the row with the majority and turnout statistics. The Object section highlights the row with the registered electors and swing percentage.

# Vertical Listings

Predicate

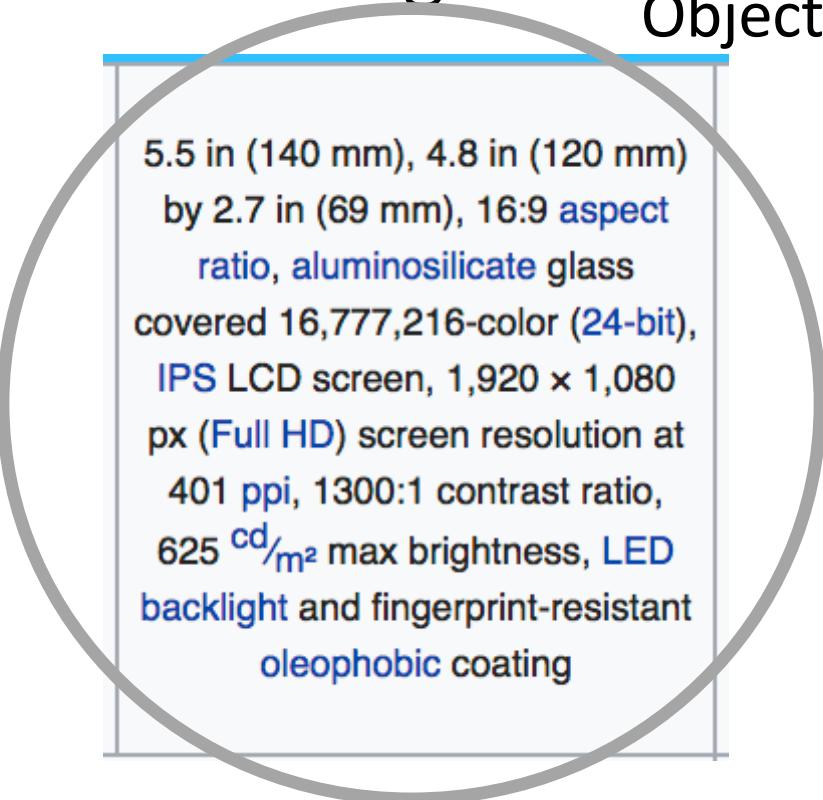
Model	[hide]	iPhone 7	iPhone 7 Plus
Picture			
Initial release operating system			iOS 10.0
Latest release operating system			
In development			
Display		4.7 in (120 mm), 4.1 in (100 mm) by 2.3 in (58 mm), 16:9 aspect ratio, aluminosilicate glass covered 16,777,216-color (24-bit), IPS LCD screen, 1,334 × 750 px screen resolution at 326 ppi, 1400:1 contrast ratio, 625 cd/m² max brightness, LED backlight and fingerprint-resistant oleophobic coating	5.5 in (140 mm), 4.8 in (120 mm) by 2.6 in (69 mm), 16:9 aspect ratio, aluminosilicate glass covered 16,777,216-color (24-bit), IPS LCD screen, 1,920 × 1,080 px (Full HD) screen resolution at 401 ppi, 1300:1 contrast ratio, 625 cd/m² max brightness, LED backlight and fingerprint-resistant oleophobic coating
Storage		32, 128, and 256 GB NAND Flash driven by NVMe Express controller	
Processor		2.33 GHz quad-core Apple-designed 64-bit Apple A10 Fusion (4-cores: 2 Hurricane high-performance, 2 Zephyr high-efficiency) with embedded M10 motion coprocessor	
Bus width			
Graphics		Custom Apple PowerVR GT7600 Plus (hexa-core) GPU <sup>[7]</sup>	
RAM		2 GB LPDDR4 DRAM	3 GB LPDDR4 DRAM

Subject

Object

# Vertical Listings

Object?



5.5 in (140 mm), 4.8 in (120 mm)  
by 2.7 in (69 mm), 16:9 aspect  
ratio, aluminosilicate glass  
covered 16,777,216-color (24-bit),  
IPS LCD screen, 1,920 × 1,080  
px (Full HD) screen resolution at  
401 ppi, 1300:1 contrast ratio,  
625 cd/m<sup>2</sup> max brightness, LED  
backlight and fingerprint-resistant  
oleophobic coating

This looks like a complex object...

Might need to generate URIs for the object entities and extract attributes from these cells

# Attribute/Value Tables

The diagram illustrates the structure of an attribute-value table. It features three main components: 'Predicate' on the left, 'Object' on the right, and a central intersection area. The 'Object' section contains a table with various attributes and their corresponding values. A green oval encircles the 'Brand' column, and a grey oval encircles the 'Apple Inc.' value, highlighting the intersection of predicate and object.

<b>Brand</b>	Apple Inc.
<b>Manufacturer</b>	Foxconn <sup>[1]</sup> (on contract) Pegatron <sup>[1]</sup> (on contract)
<b>Slogan</b>	Say hello to the future
<b>Generation</b>	11th
<b>Model</b>	A1865 (with Qualcomm modem) A1901 (with Intel modem) A1902 (sold in Japan) <sup>[2]</sup>
<b>Compatible networks</b>	GSM, CDMA2000, EV-DO, HSPA+, LTE, LTE Advanced
<b>First released</b>	November 3, 2017
<b>Availability by Region</b>	November 3, 2017 November 23, 2017 November 24, 2017 December 1, 2017 December 7, 2017 December 8, 2017 December 22, 2017
<b>Discontinued</b>	September 12, 2018
<b>Predecessor</b>	iPhone 7 / iPhone 7 Plus
<b>Successor</b>	iPhone XS / iPhone XS Max iPhone XR
<b>Type</b>	Smartphone
<b>Form factor</b>	Slate
<b>Dimensions</b>	H: 143.6 mm (5.65 in) W: 70.9 mm (2.79 in) D: 7.7 mm (0.30 in)

Where's the subject???

# Attribute/Value Tables

Subject in  
Attribute/Value tables is  
often outside the table

Predicate

iPhone X		
iPhone X		
Brand	Apple Inc.	
Manufacturer	Foxconn <sup>[1]</sup> (on contract) Pegatron <sup>[2]</sup> (on contract)	
Slogan	Say hello to the future	
Generation	11th	
Model	A1865 (with Qualcomm modem) A1901 (with Intel modem) A1902 (sold in Japan) <sup>[2]</sup>	
Compatible networks	GSM, CDMA2000, EV-DO, HSPA+, LTE, LTE Advanced	
First released	November 3, 2017	
Availability by Region	November 3, 2017 [show] November 23, 2017 [show] November 24, 2017 [show] December 1, 2017 [show] December 7, 2017 [show] December 8, 2017 [show] December 22, 2017 [show]	

Subject

Object

Matrix

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1920	469	391	693	761	627	723	553	790	777	796	823	
1921	743	794	750	748	874	586	538	885	881	747	869	525
1922	727	583	806	924	771	826	893	764	1,127	741	879	1,122
1923	762	666	1,028	1,511	1,324	2,598	1,547	1,542	1,592	1,315	1,397	2,103
1924	1,528	1,680	1,550	1,547	1,858	1,542	1,409	1,242	1,893	1,488	1,453	1,049
1925	1,149	1,122	1,058	798	1,356	1,255	1,302	1,465	923	1,292	740	877
1926	1,183	1,049	965	1,308	1,842	1,226	1,726	1,083	1,388	1,010	1,344	1,283
1927	1,204	1,165	1,199	1,171	1,390	1,411	1,089	1,382	1,297	1,539	1,280	1,717
1928	1,225	1,243	1,530	1,303	1,493	1,879	1,669	1,883	1,506	2,015	1,691	1,529
1929	1,972	1,678	1,600	1,726	1,932	2,615	3,117	2,236	1,988	2,869	2,579	2,089
1930	1,808	1,731	1,944	1,900	2,202	2,508	1,973	2,407	1,961	2,167	1,765	1,339
1931	1,919	1,710	1,586	1,826	2,268	2,544	2,621	2,856	2,296	2,389	2,449	1,071
1932	1,592	1,897	2,090	2,867	2,942	2,791	2,249	2,839	2,113	2,541	1,318	2,154
1933	1,913	1,886	2,137	2,939	2,679	4,355	4,523	3,141	3,182	3,888	3,305	2,636
1934	2,288	2,511	2,582	3,942	3,724	3,794	4,128	3,696	4,068	3,277	4,680	2,437
1935	2,369	2,804	3,281	3,776	4,613	5,589	5,832	4,946	4,971	4,810	4,289	4,098
1936	3,067	3,474	3,155	3,743	4,390	4,792	4,458	5,561	5,025	4,708	4,145	3,666
1937	3,596	3,777	3,196	4,899	6,796	6,181	6,363	7,423	6,602	6,692	6,645	5,116
1938	5,953	5,328	6,121	7,553	7,798	7,424	7,250	7,003	5,577	6,780	5,602	4,884
1939	4.477	4.810	4.966	6.222	8.643	5.831	7.304	5.969	6.925	6.947	5.323	4.656

# Matrix - Calendar

SUNDAY	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY
30  Mostly Sunny Actual: 76°   61° 0 in	1  Partly Cloudy Actual: 83°   66° 0 in	2  Partly Sunny Actual: 76°   67° 0 in	3  Rain Actual: 77°   67° 0 in	4  Partly Cloudy Actual: 75°   65° 0 in	5  Partly Cloudy Actual: 73°   64° 0 in	6  Partly Sunny Actual: 75°   64° 0 in
7  Partly Cloudy Actual: 74°   61° 0 in	8  Partly Cloudy Actual: 73°   64° 0 in	9  Partly Cloudy Actual: 74°   64° 0 in	10  Partly Cloudy Actual: 72°   62° 0 in	11  Mostly Sunny Forecast: 74°   61° 0 in	12  Sunny Forecast: 76°   64° 0.01 in	13  Mostly Cloudy Forecast: 68°   61° 0 in
14  Partly Cloudy Forecast: 68°   60° 0 in	15  Sunny Forecast: 77°   57° 0 in	16  Sunny Forecast: 76°   59° 0 in	17  Sunny Forecast: 78°   59° 0 in	18  Sunny Forecast: 78°   60° 0 in	19  Sunny Forecast: 78°   61° 0 in	20  Sunny Forecast: 78°   62° 0 in
21  Sunny Forecast: 77°   63° 0 in	22  Sunny Forecast: 76°   64° 0 in	23  Sunny Forecast: 75°   63° 0 in	24  Sunny Forecast: 74°   63° 0 in	25 Average: 72°   58° 0 in	26 Average: 72°   58° 0 in	27 Average: 72°   57° 0 in
28 Average: 72°   57° 0 in	29 Average: 72°   57° 0 in	30 Average: 72°   57° 0 in	31 Average: 72°   57° 0 in	1 Average: 71°   56° 0 in	2 Average: 71°   56° 0 in	3

Subject

Object

Enumeration      Predicate

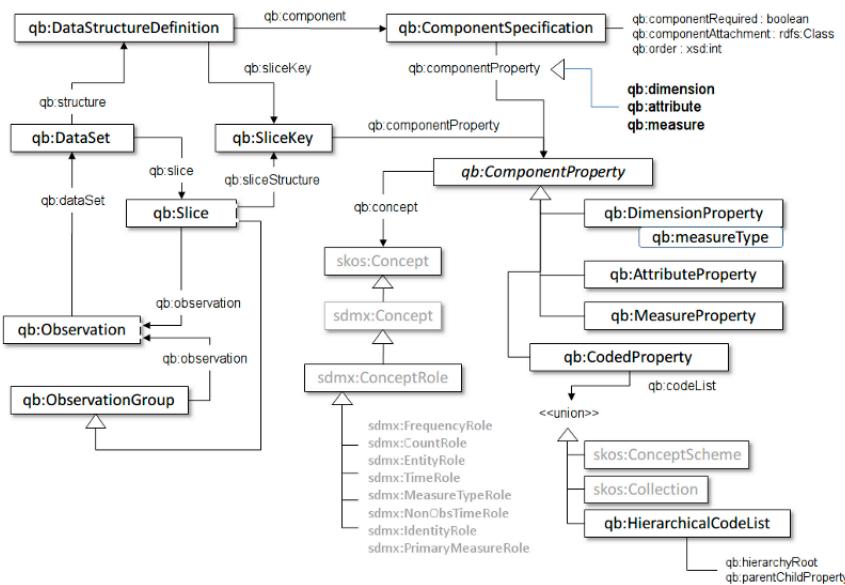
Classes in the OWL vocabulary:	Subject
<a href="#">rdfs:Class</a>	
<a href="#">owl:AllDifferent</a>	
<a href="#">owl:AnnotationProperty</a>	
<a href="#">owl:Class</a>	
<a href="#">owl:DataRange</a>	
<a href="#">owl:DatatypeProperty</a>	
<a href="#">owl:DeprecatedClass</a>	
<a href="#">owl:DeprecatedProperty</a>	
<a href="#">owl:FunctionalProperty</a>	
<a href="#">owl:InverseFunctionalProperty</a>	
<a href="#">owl:Nothing</a>	
<a href="#">owl:ObjectProperty</a>	
<a href="#">owl:Ontology</a>	
<a href="#">owl:OntologyProperty</a>	
<a href="#">owl:Restriction</a>	
<a href="#">owl:SymmetricProperty</a>	
<a href="#">owl:Thing</a>	
<a href="#">owl:TransitiveProperty</a>	

# Extracting Quantitative Data from Tables

# Representing Quantitative Data

- What is quantitative data?
- How is it harder than relational data?
- How is it easier than relational data?
- What is important to represent about this data?

# RDF Data Cube Ontology



# Key ideas in Quantitative Data

- Measure:
  - the thing we want to measure
- Attributes:
  - help us understand what we're measuring
- Dimensions:
  - values that are varied across measurements
- Observation:
  - The actual value we measured

# Key ideas in Quantitative Data: Example

- Measure: Temperature
  - the thing we want to measure
- Attributes: Time of Day, Elevation, Unit
  - help us understand what we're measuring
- Dimensions: Location
  - values that are varied across measurements
- Observation: 70° F
  - The actual value we measured

# Key ideas in Quantitative Data: Example

- Measure: Temperature
  - the thing we want to measure
- Attributes: Location, Unit
  - help us understand what we're measuring
- Dimensions: Time, Elevation
  - values that are varied across measurements
- Observation: 70° F
  - The actual value we measured

# Key ideas in Quantitative Data: Example

- Measure: INF-558 Grades
  - the thing we want to measure
- Attributes:
  - help us understand what we're measuring
- Dimensions:
  - values that are varied across measurements
- Observation:
  - The actual value we measured

# Challenges in Extracting Quantitative Data

CARE FOOD PROGRAM				
	Meals Served			
State or District	Total	Homes	Adult	Centers
Alabama	1062	542	509	11
Alaska	866	414	448	4
District of Columbia	792	357	433	2

FT 2011 data are preliminary; all data are subject to revision.  
RP = Reduced Price

(a)

CARE FOOD PROGRAM				
	Meals Served			
State or District	Total	Homes	Adult	Centers
Alabama	1062	542	509	11
Alaska	866	414	448	4
District of Columbia	792	357	433	2

FT 2011 data are preliminary; all data are subject to revision.  
RP = Reduced Price

(b)

Filling in Implicit Values			
Attribut A	Attribut B	Attribut C	Attribut D
A1	B1	C1	D1
		C2	D2
B2		C3	
		C4	D3
B3	C5		D4

(c)

- What are the attributes?
  - Unit detection?
  - Footnotes and other metadata
- What are the dimensions?
- What are the values?
  - How are unknown values indicated?
  - Are some values estimated or predicted?
  - Are some values aggregates?

Figure courtesy DeExcelerator: A Framework for Extracting Relational Data From Partially Structured Documents (Eberius et al., CIKM13)

# Challenges in Quantitative Data

	Consumer Price Inflation (1)	General Government Surplus (+) or Deficit (-)	General Government Gross Debt	Long-Term Interest Rate	Exchange Rates Within the EMS		All Criteria Fulfilled	
	% change	% of GDP	% of GDP	%	1995/96	1996	„Strict“ Interpretation (2)	Art. 104c(2) Interpretation (2)
Belgium	1.9	-2.7	126.7	5.8	yes	yes	no	yes
Denmark	2.4	+0.3	67.2	6.2	yes	yes	no (4)	yes (5)
Germany	1.8	-3.0	61.8	5.6	yes	yes	no	yes

topology: short-cuts?	slopes: conventional mean $\pm$ st.d. quartiles	centered		
		direct comparison: # of faster runs	mean $\pm$ st.d. quartiles	*
no	<b>65.4 <math>\pm</math> 15.9</b> 57/62/70	52 – 48	<b>51.6 <math>\pm</math> 16.2</b> 43/64.5/ $\infty$	
	81 17	0 $\times$ 61 39 99	4 95	
yes	<b>90.4 <math>\pm</math> 31.1</b> 69.5/80/102	0 – 100	<b>33.1 <math>\pm</math> 8.6</b> 28/31/35	

From Table Header Detection and Classification, (Fang, Mitra, Zang, Giles, AAAI12)

# Challenges in quantitative datasets

Data spec in a cell next

to months

Part I Selected monthly macroeconomic indicators		Data update date: 24.07.2018				
		2009				
Specification		IV	V	VI	VII	VIII
A – corresponding period of the previous year=100						
A <sub>t</sub> – from the beginning of the year to the end of the period (corresponding period of the previous year=100)						
B – the previous period=100						
C – December of the previous year=100						
I <sub>2</sub> – monthly average of 2005=100						
I <sub>3</sub> – monthly average of 2010=100						
monthly average of 2010=100						
Average paid employment in an enterprise sector <sup>a</sup>	in thous.	5,309	5,292	5,280	5,273	5,270
A		98.6	98.3	98.1	97.8	97.8
B		99.7	99.7	99.8	99.9	99.9
I <sub>2</sub>		111.0	110.7	110.5	110.4	110.3
I <sub>3</sub>						
Registered unemployed persons (end of the period)	in thous.	1,719.9	1,683.4	1,658.7	1,676.1	1,689.0
A		107.1	110.3	114.0	117.8	120.3
B		97.8	97.9	98.5	101.1	100.8

DATA YEARS	Ludność Population		Małżeństwa Marriages	Rozwody Divorces	Urodzenia Live births	Zgonы Deaths		Migracje wewnętrzne Internal migration			Migracje zagraniczne <sup>c</sup> International migration <sup>c</sup>			Ogólne saldo migracji Total net migration	Małżeństwa Marriages	Rozwody Divorces	Urodzenia Live births
	stan w dniu 30 VI as of June 30	stan w dniu 31 XII as of De- cember 31				w tym ogółem total	nitemówka of which infant	Przyrost naturalny Natural increase	napływ inflow	odpływ outflow	saldo net	imigracja immigra- tion	emigracja emigra- tion	saldo net			
	w tysiącach	in thousands															

Holy dimensional  
nesting, Batman!

Table  
segmentation

Months in  
Roman  
numerals

URBAN AREAS		
Age groups	1989	1990
TOTAL	23384	23546
0 - 2 years	964	920
3 - 6	1571	1523
7 - 14	3090	3054
15 - 18	1415	1446
2 of which: 18	332	342
19 - 24	1829	1897
4 - 17	6708	6600
5 18 - 59/64	14020	14209
6 of which:		
7 18 - 44	9787	9838
8 45 - 59/64	4233	4371
9 60/65 years and more	2656	2737
0 0 - 14	5626	5496
15 - 64	15689	15904
65 years and more	2070	2147
3 Women at childbearing 15 - 49 years	6182	6255
4		
6 TOTAL	x	162.0
7 0-2 years	x	-44.0
8 3 - 6	x	-48.0
9 7 - 14	x	-36.0
0 15 - 18	x	31.0
1 of which: 18	x	10.0
19 - 24	x	68.0
3 0 - 17	x	-108.0
4 18 - 59/64	x	189.0
5 of which: 18		
6 18 - 44	y	51.0

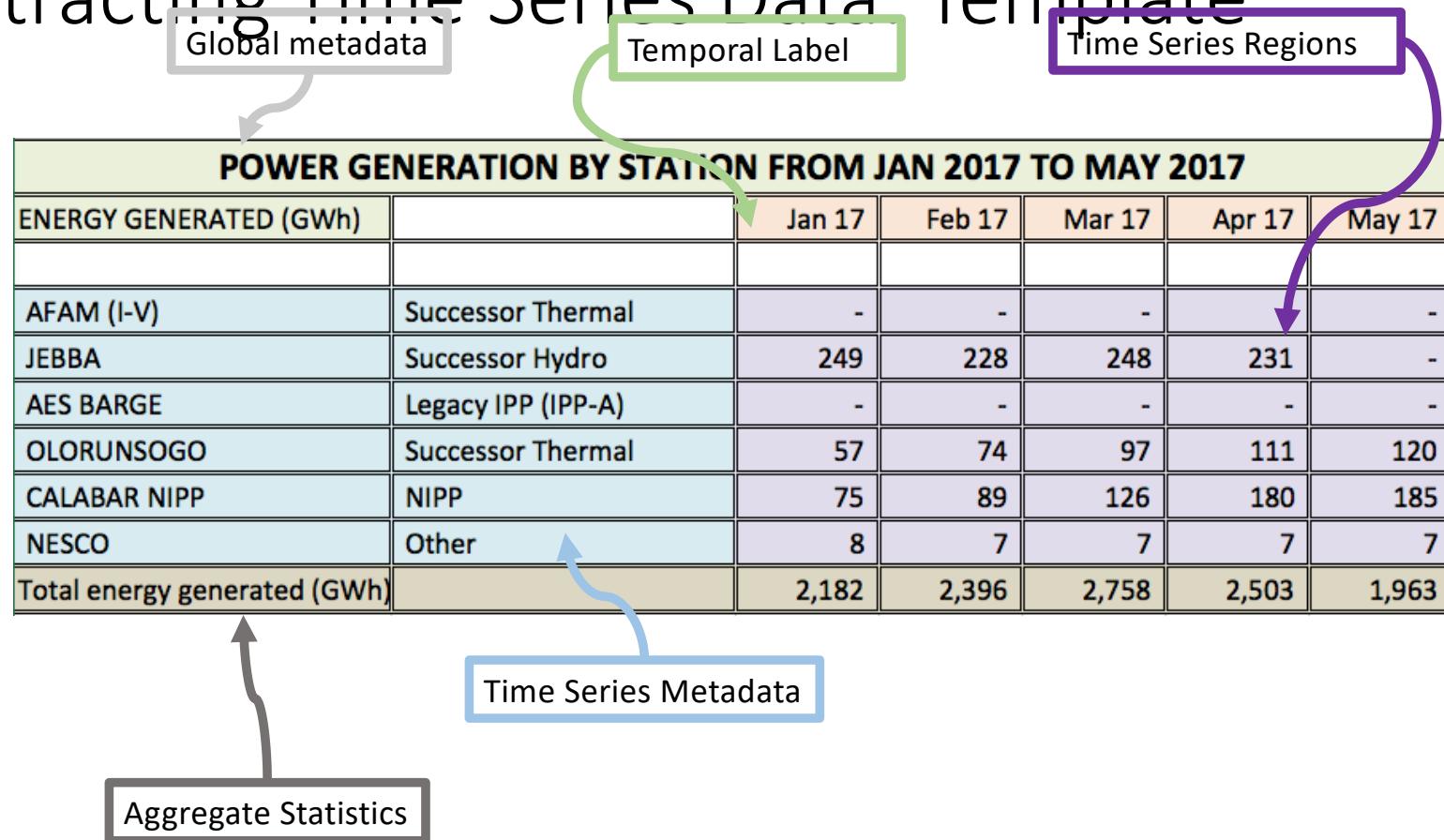
# Approaches to structured data extraction

- Table Segmentation
- Cell classification
  - String, Date, Number, Blank
- Layout detection
  - Remove formatting/padding
  - Nested headers, implicit or inherited values
- Table Specifications
  - Metadata, Dimension, Observation
  - Derived or Aggregate cells
- Entity Linking, Semantic Labeling
  - Type detection
  - Data Augmentation Approaches
- Extraction into triples

# Motivating Example: (Simple) Time Series

- Tabular datasets
- One dimension is always time
- Other dimensions are orthogonal to time
- Values share a single type

# Extracting Time Series Data: Template



# Extracting Time Series: Annotations

			E	F	G	H	
1	POWER GENERATION BY STATION FROM JAN 2017 TO MAY 2017						
2	ENERGY GENERATED (GWh)		Jan 17	Feb 17	Mar 17	Apr 17	May 17
3							
4	AFAM (I-V)	Successor Thermal	-	-	-	-	-
11	JEBBA	Successor Hydro	249	228	248	231	-
12	AES BARGE	Legacy IPP (IPP-A)	-	-	-	-	-
19	OLORUNSOGO	Successor Thermal	57	74	97	111	120
27	CALABAR NIPP	NIPP	75	89	126	180	185
28	NESCO	Other	8	7	7	7	7
29	Total energy generated		82	2,396	2,758	2,503	1,963

```
"metadata": [
    { "loc": "B", "name": "label"}, 
    { "loc": "C", "name": "station_type"}]
```

```
"orientation": "row",
"times": { "locs": "[2]" }
```

```
"Metadata": [
    { "name" : "description",
      "source" : "cell",
      "loc" : "(A,1)" },
    { "name" : "title",
      "source" : "sheet_name" }]
```

```
"rows": "[4:25]",
"locs": "[D:H]"
```

# Abstracting TS template to specification

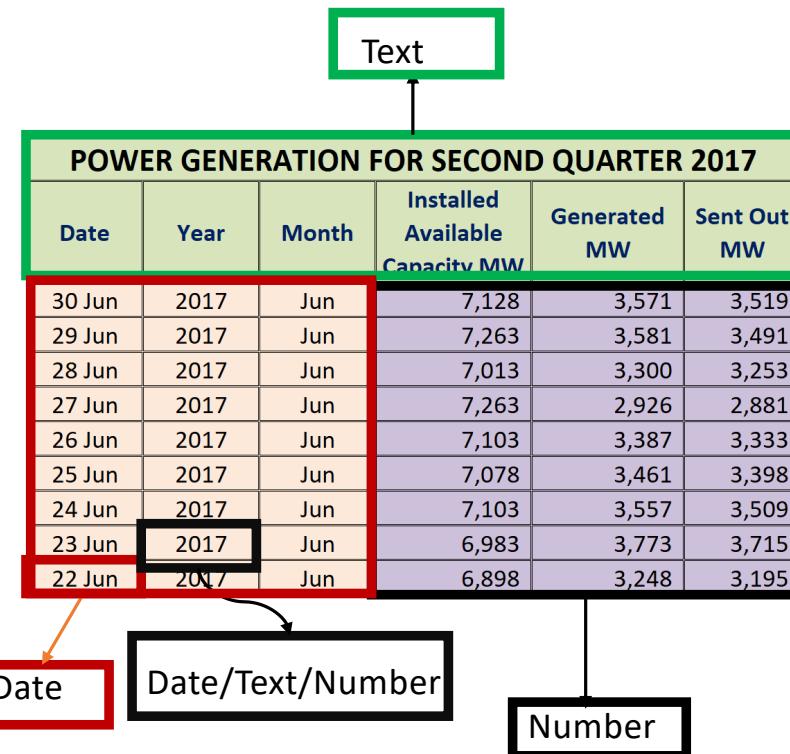
```
"TimeSeriesRegions":
```

```
[  
  {  
    "orientation": "row",  
    "rows": "[4:25]",  
    "locs": "[D:H]",  
    "times": { "locs": "[2]" },  
    "metadata": [  
      { "loc": "B", "name": "label" },  
      { "loc": "C", "name": "station_type" },  
    ],  
  },  
,  
  "Metadata": [  
    { "name" : "description",  
      "source" : "cell",  
      "loc" : "(A,1)"  
    },  
    { "name" : "title",  
      "source" : "sheet_name"  
    }  
]
```

	B	C	D	E	F	G	H
1	POWER GENERATION BY STATION FROM JAN 2017 TO MAY 2017						
2	ENERGY GENERATED (GWh)		Jan 17	Feb 17	Mar 17	Apr 17	May 17
3							
4	AFAM (I-V)	Successor Thermal	-	-	-	-	-
11	JEBBA	Successor Hydro	249	228	248	231	-
12	AES BARGE	Legacy IPP (IPP-A)	-	-	-	-	-
19	OLORUNSOGO	Successor Thermal	57	74	97	111	120
27	CALABAR NIPP	NIPP	75	89	126	180	185
28	NESCO	Other	8	7	7	7	7
29	Total energy generated (GWh)		2,182	2,396	2,758	2,503	1,963

# Automated Annotation

- Classify cells
- Detect rectangular time blocks
- Validate the detected block
- Orientation
- Parse date
- Granularity
- Labels



Dissect complicated tables

# A unified framework for table understanding

Pujara et al., ISWC 2019

# Tables contain really useful data

**Table .1 Estimates of Cropland Area, Expected Production and Yield of Major Crops Forecast For Private Peasant holdings: 2018/19 (2011 E.C.), Meher Season**

**Table .2 Estimates of Cropland Area, Expected Production and Yield of Major Crops Forecast For Private Peasant holdings; 2018/19 (2011 E.C.), Meher Season**

# Vast Amount of Data in Tabular Form

Socio-political



Financial and business



Health and environment

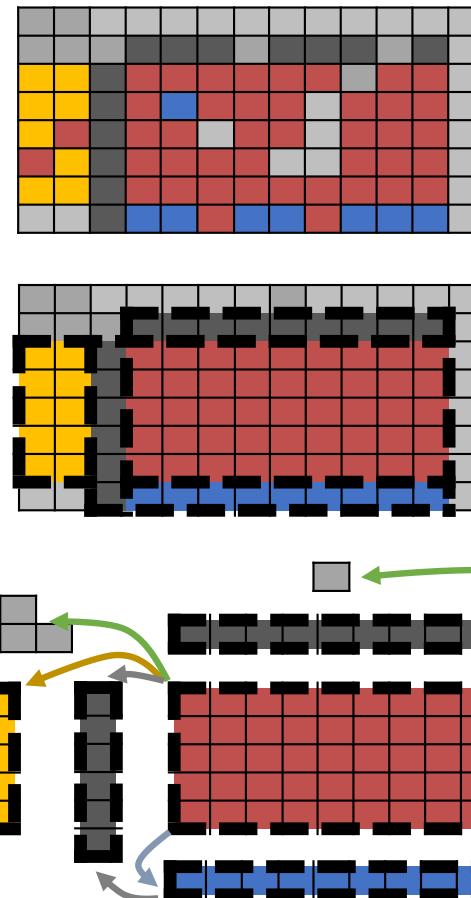
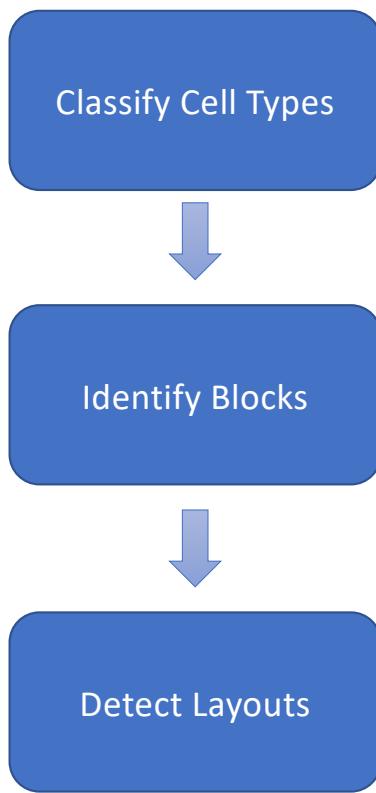


# Tables contain really useful data

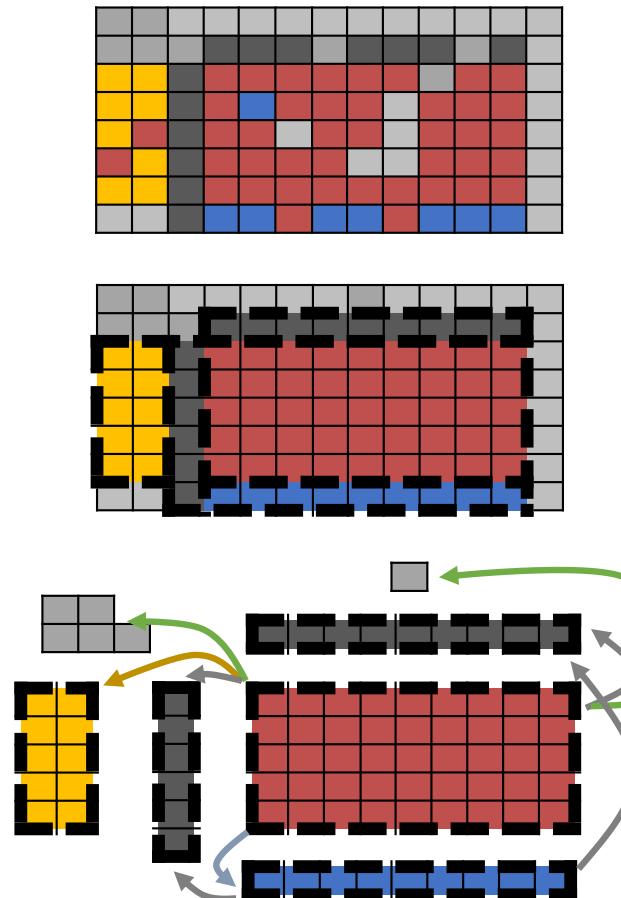
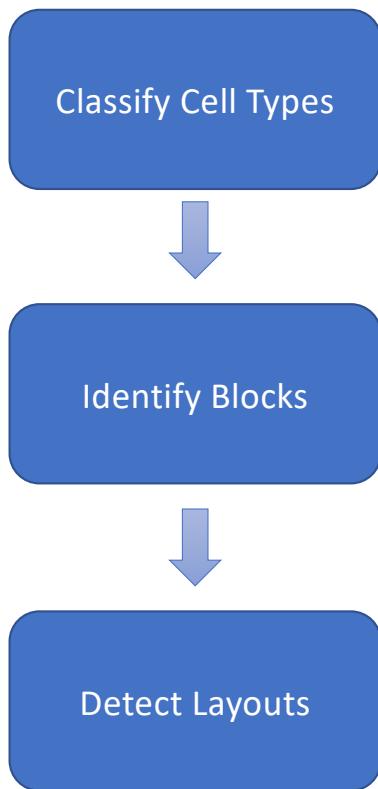
**Table .1 Estimates of Cropland Area, Expected Production and Yield of Major Crops Forecast For Private Peasant holdings; 2018/19 (2011 E.C.), Meher Season**

# How do you extract useful data from tables?

# Three core tasks across all table understanding tools

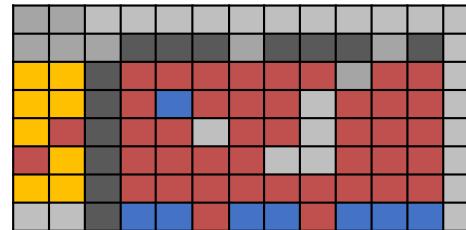
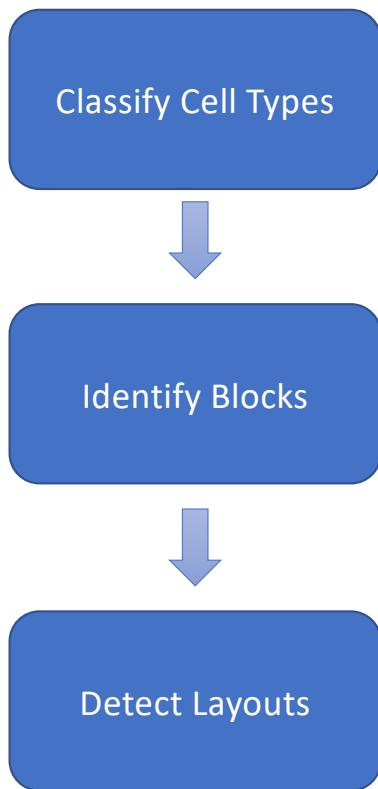


# Three core tasks across all table understanding tools

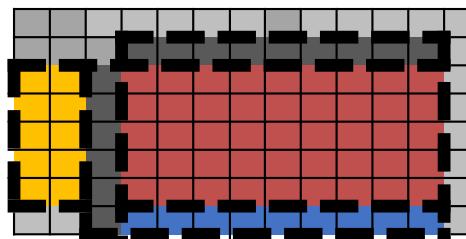


Syntactic	Semantic
datatypes (string, date)	semantic type (Person, Time)
role-based (header)	ontology- based (Companies)
relational joins (indexing, hierarchy)	semantic model (properties, isA)

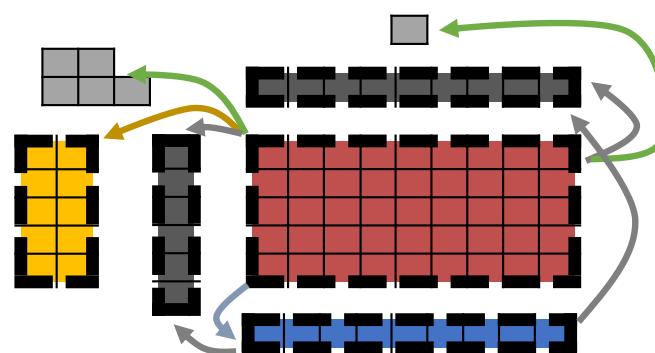
# Release of framework at <https://github.com/usc-isi-i2/isi-table-understanding>



conditional random fields



Entropy-based partitioning



conditional random fields

# Utilities for interoperability

Flexible and expressive output

- Colorized Excel easy for humans to diagnose
- YAML/JSON outputs for reusable data models
- Normalized dataframes for downstream tasks

State	County	Health Outcomes		Health Factors	
		Z-Score	Rank	Z-Score	Rank
Alabama					
Alabama	Autauga	-1.07	7	-0.49	11
Alabama	Baldwin	-1.54	2	-0.86	3
Alabama	Barbour	-0.17	33	0.56	58
Alabama	Bibb	-0.01	40	-0.12	29

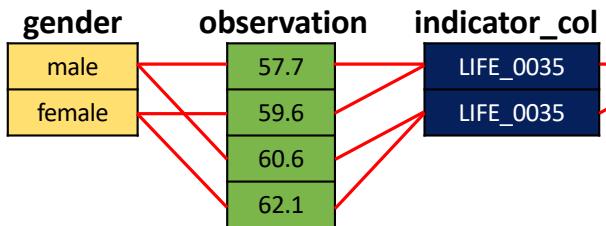
# D-REPR language

(Vu et al, KCAP 2019)

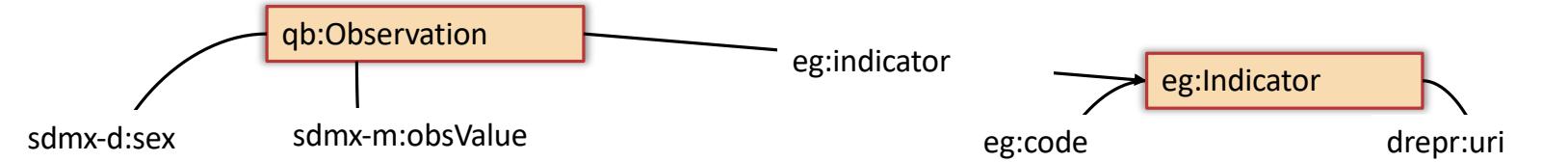
```
{
  "indicator": "LIFE_0035",
  "url": "http://apps.who.int/.../indicator.aspx?iid=35"
},
{
  "indicator": "LIFE_0029",
  "url": "http://apps.who/int/.../indicator.aspx?iid=29"
},
```

## Step 1: define resources

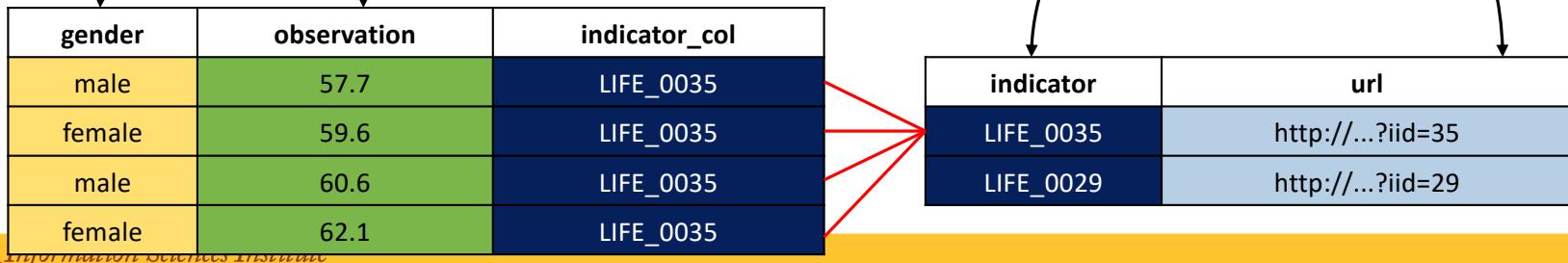
		2016	
Indicator	Age Group	Male	Female
LIFE_0035	<1 year	57.7	59.6
LIFE_0035	1-4 years	60.6	62.1



## Step 2: define blocks



## Step 4: semantic modeling



## Step 3: join attributes to tables

# Easy-to-use interfaces

ML-ready with probabilistic, hierarchical outputs:

- Output probability distributions
- Capture different levels of granularity by introducing parent-child relationships
- Mix-and-match components

```
from reader.sheet import Sheet
from type.cell.cell_type_pmf import CellTypePMF
from typing import List

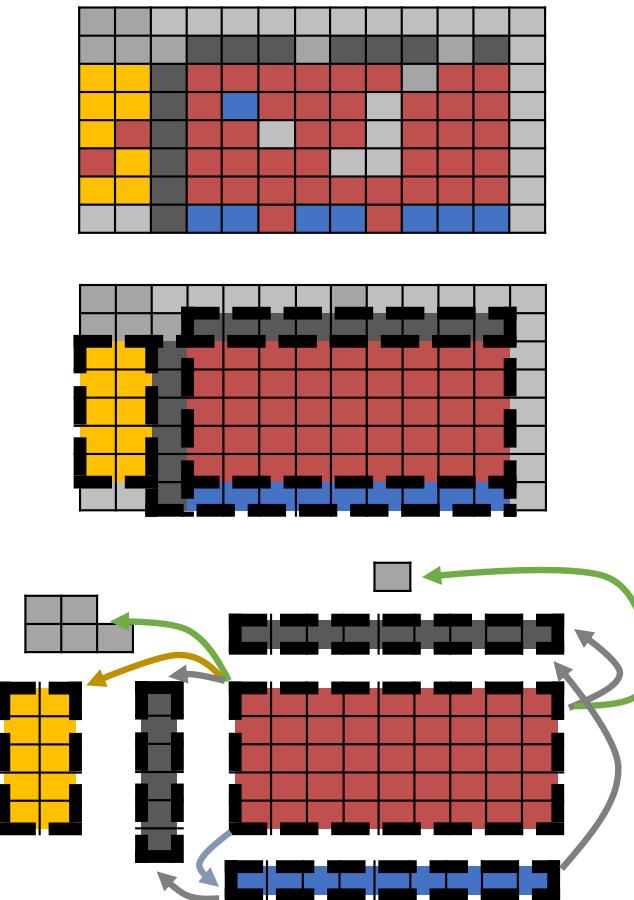
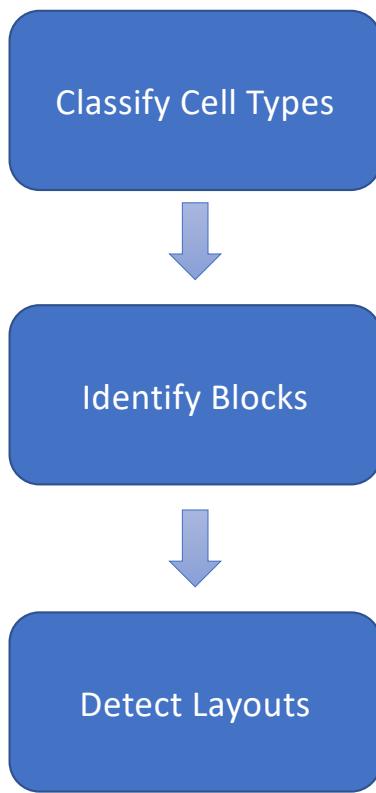
class CellClassifier(abc.ABC):
    @abc.abstractmethod
    def classify_cells(self, sheet: Sheet) -> 'np.ndarray[CellTypePMF]':
        pass
```

Dissect complicated tables

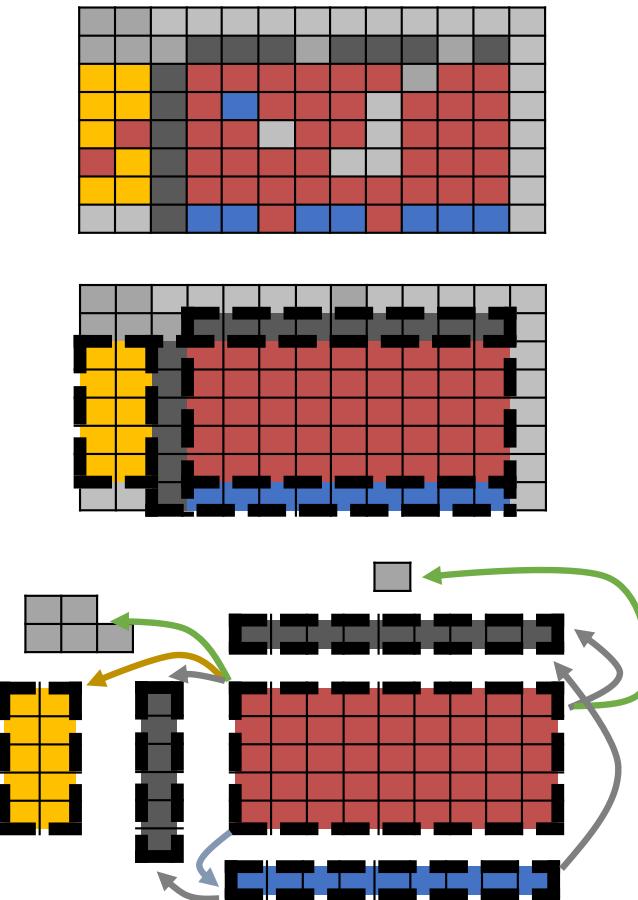
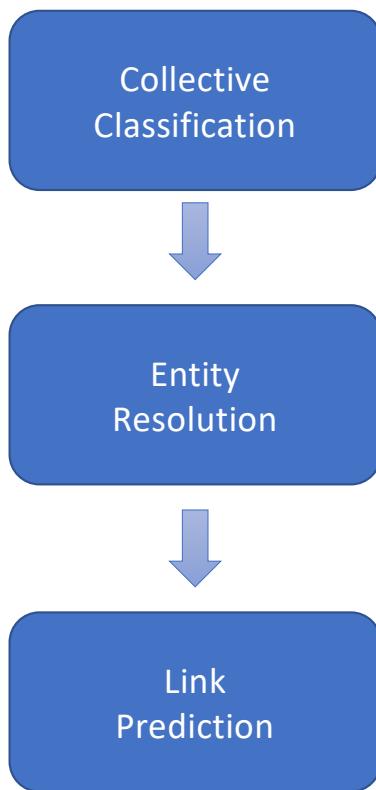
# A probabilistic framework for table understanding

Kexuan Sun and Harsha Rayudu

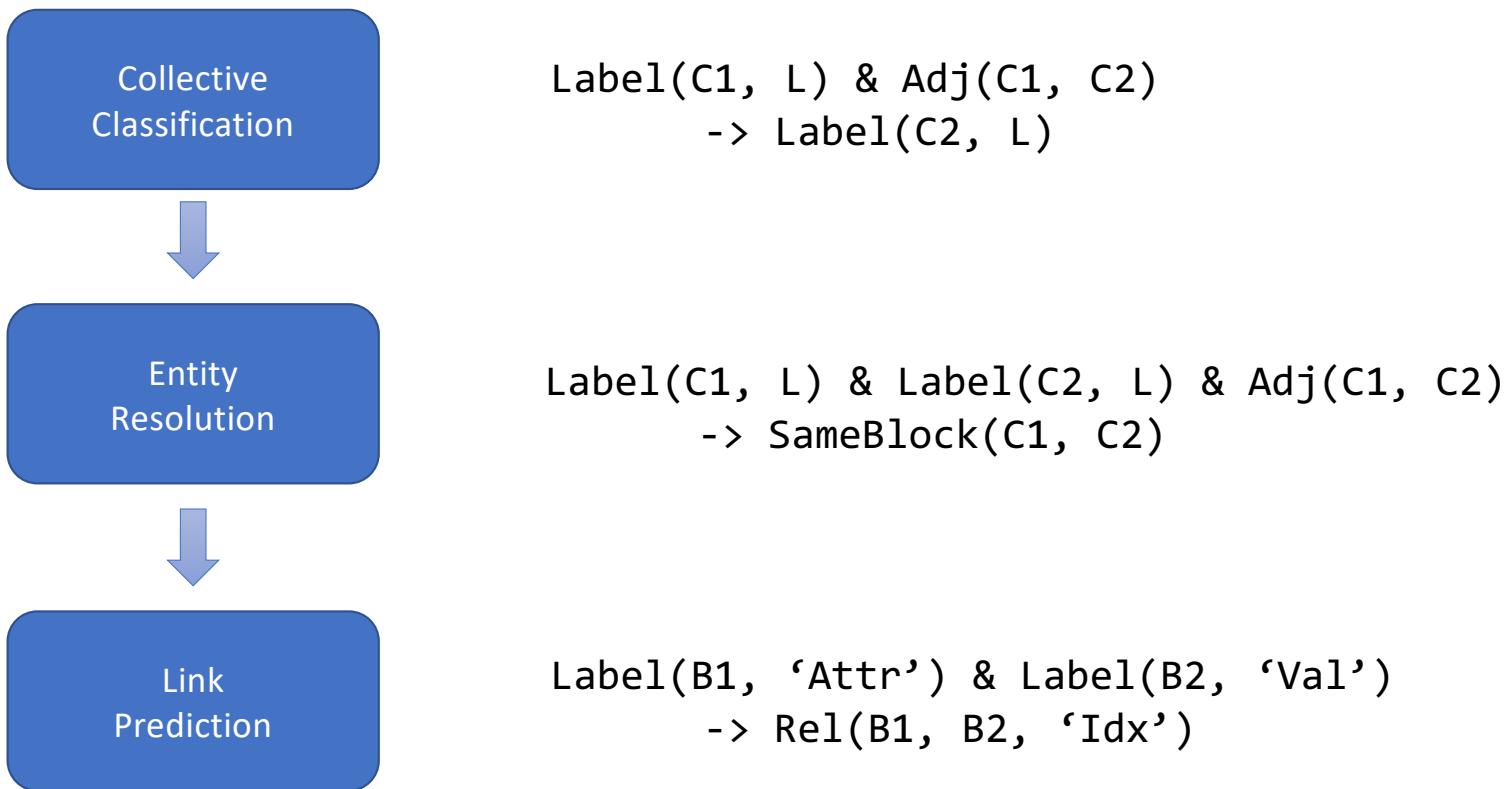
# Three core tasks across all table understanding tools



# Map nicely to tasks in statistical relational learning



# Modeling probabilistic relationships (via PSL)



Dissect complicated tables

# Pre-trained tabular cell embeddings

Gol et al., ICDM 2019

# Why Cell Embeddings for Tabular Data?

Table1 - 2018-19 academic year course evaluations								
Computer Science Department								
	Course Information					Course Evaluation*		Semester Average Evaluation Score
	Course Code	Course Name	Instructor Name	Number of Registered Students	Start Date	Exam Date	Number of Students Responded to course Evaluation	
Fall 2018	CSCI561	Foundations of Artificial Intelligence	Prof. Tejada	540	8/26/18	12/2/18	45	4.8
	CSCI567	Machine Learning	Prof. Sha	334	8/27/18	12/5/18	32	4.9
	CSCI585	Database Systems	Prof. Shahabi	460	8/30/18	12/4/18	44	4.9
Spring 2019	CSCI561	Foundations of Artificial Intelligence	Prof. Tejada	540	1/9/19	5/2/19	75	4.9
	CSCI670	Advanced Analysis of Algorithms	Prof. Kempe	70	1/11/19	5/7/19	12	5

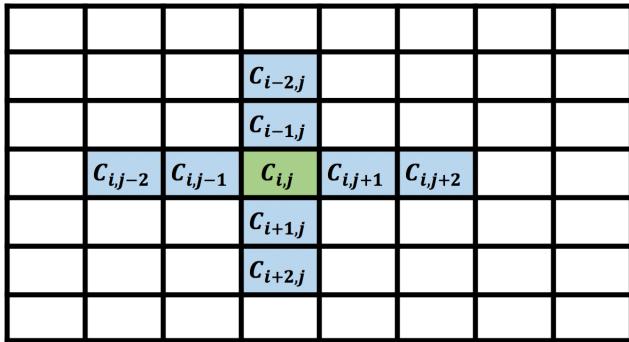
*\*Evaluation surveys were completed before the final exam date. Survey links had been emailed to all students.*

[Wang 1996]

- Tabular data contains multi-dimensional relations
- Tabular data is designed for easy human consumption
  - Order of data dimensions, and values often follow some rules
  - Cell styling often provides visual clues
- Tabular data **follows regular patterns in data formation (layout)**

# Proposed Cell Embedding Approach

**Goal:** given a corpus of tabular documents  $D = \{C_{i,j}; 1 \leq i \leq N, 1 \leq j \leq M\}$ , learn an embedding operator ( $E$ ) that maps a cell  $C_{i,j}$  to a  $k$ -dimensional vector,  $V_{i,j} \in \mathbb{R}^k$ .



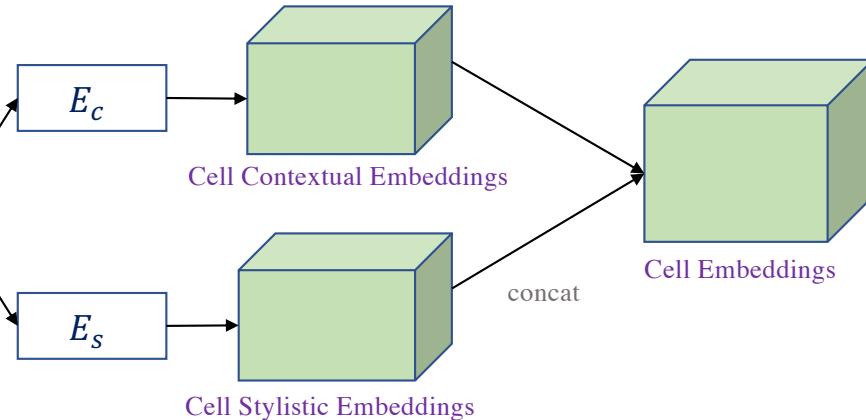
Part 1  $E_c$ : Vectorize local context of  $C_{i,j}$

Part 2  $E_s$ : Vectorize stylistic properties of  $C_{i,j}$

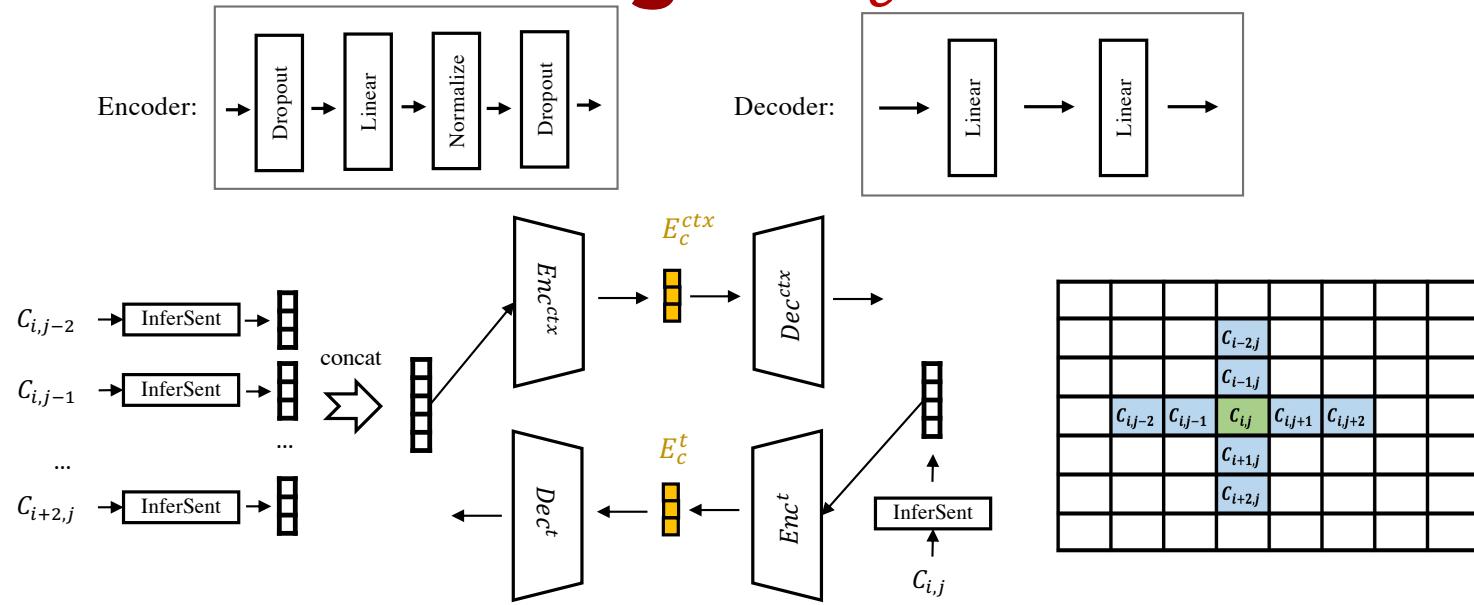
$$E(C_{i,j}) = \langle E_c(C_{i,j}), E_s(C_{i,j}) \rangle$$

Table 1 - 2018-19 academic year course evaluations							
Computer Science Department							
Course Information			Course Evaluation*			Semester Average Evaluation Score	
Course Code	Course Name	Instructor Name	Number of Registered Students	Start Date	Exam Date	Number of Students Responded to course Evaluation	Evaluation Score
Full 2018	CSC1561 Foundations of Artificial Intelligence	Prof. Tejada	540	8/26/18	12/2/18	45	4.8
	CSC1567 Machine Learning	Prof. Sha	334	8/27/18	12/5/18	32	4.9
	CSC1585 Database Systems	Prof. Shahabi	460	8/30/18	12/4/18	44	4.9
Spring 2019	CSC1561 Foundations of Artificial Intelligence	Prof. Tejada	540	1/9/19	5/2/19	75	4.9
	CSC1670 Advanced Analysis of Algorithms	Prof. Kempe	70	1/11/19	5/7/19	12	5.0

\*Evaluation surveys were completed before the final exam date. Survey links had been emailed to all students.



# Unsupervised Learning of $E_c$



$$l(\phi) = \sum_i \left| I(C_i) - Dec_{\phi_1} \left( Enc_{\phi_2}(X_{C_i}) \right) \right|^2 + \\ \sum_i \sum_{C_j \in X_{C_i}} \left| I(C_i) - Dec_{\phi_3} \left( Enc_{\phi_4}(C_i) \right) \right|^2$$

$$E_c = \underset{\Phi}{\operatorname{argmin}} l(\varphi)$$

## Loss Function

sum of mean squared error for  
all the cells in all training documents

# Model Parameters Optimization

# Unsupervised Learning of $E_S$

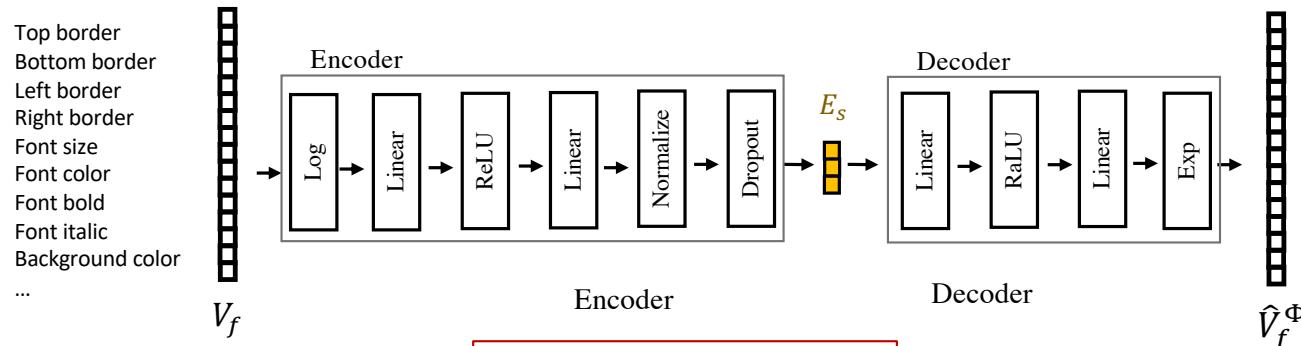
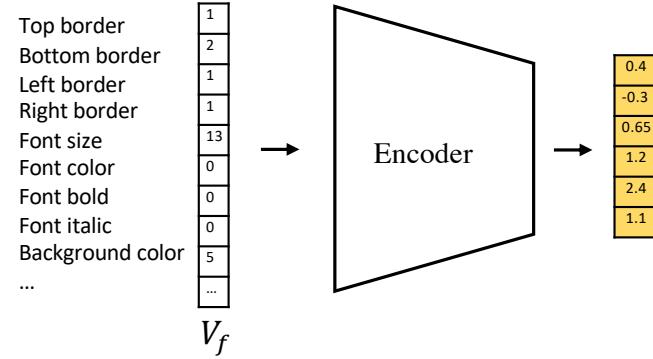


Table 1 - 2018-19 academic year course evaluations								
Computer Science Department								
LA	Course Information		TA					
	Course Code	Course Name	Instructor Name	Number of Registered Students	Start Date	Exam Date	Number of Students Responded to course Evaluation	Semester Average Evaluation Score
Fall 2018	CSCI561	Foundations of Artificial Intelligence	Prof. Tejada	546	8/26/18	12/2/18	45	4.8
	CSCI567	Machine Learning	Prof. Sha	334	8/27/18	12/5/18	32	4.9
Spring 2019							4.9	5.0
							5	

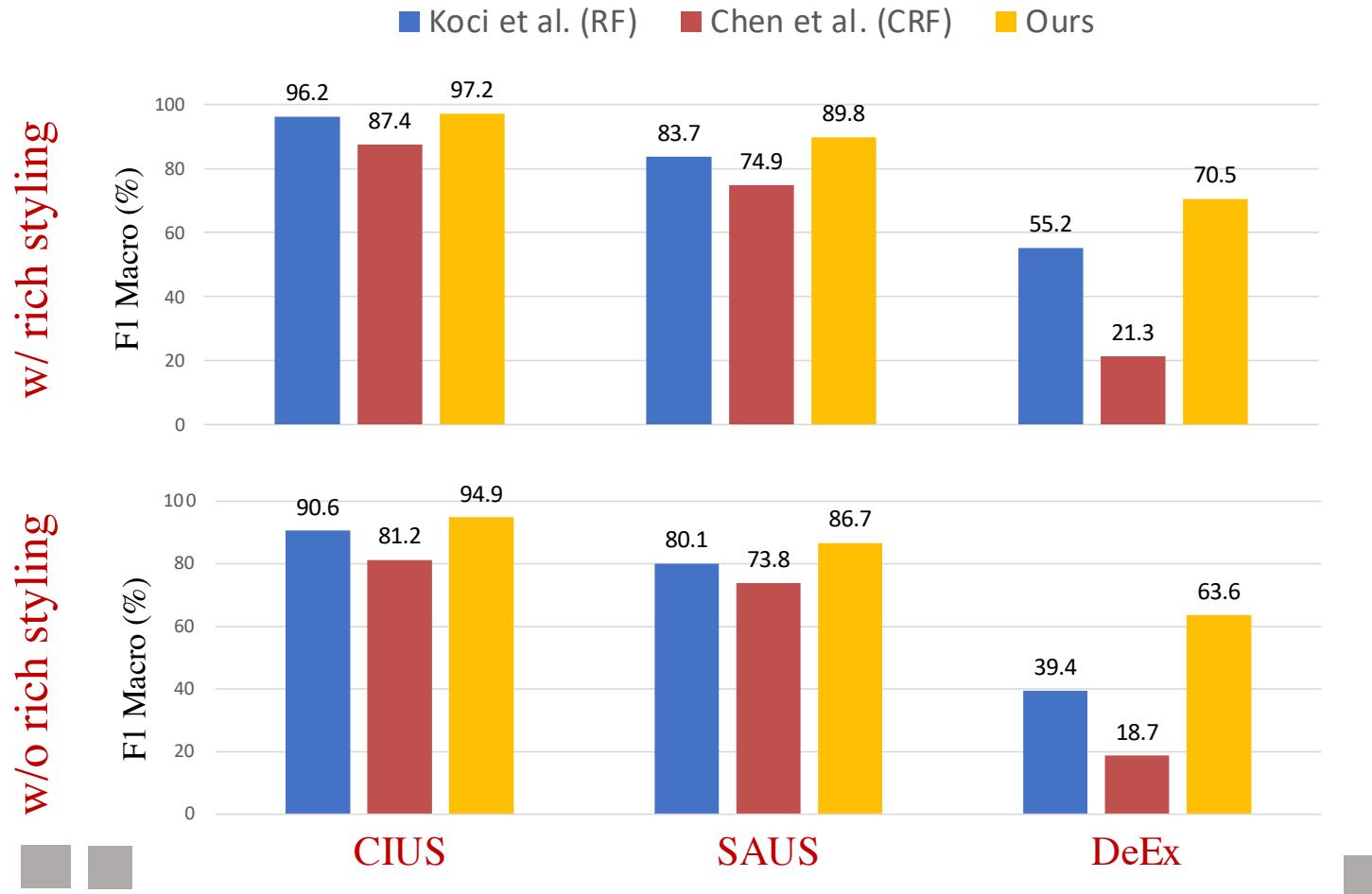
\*Evaluation score is calculated based on student responses.

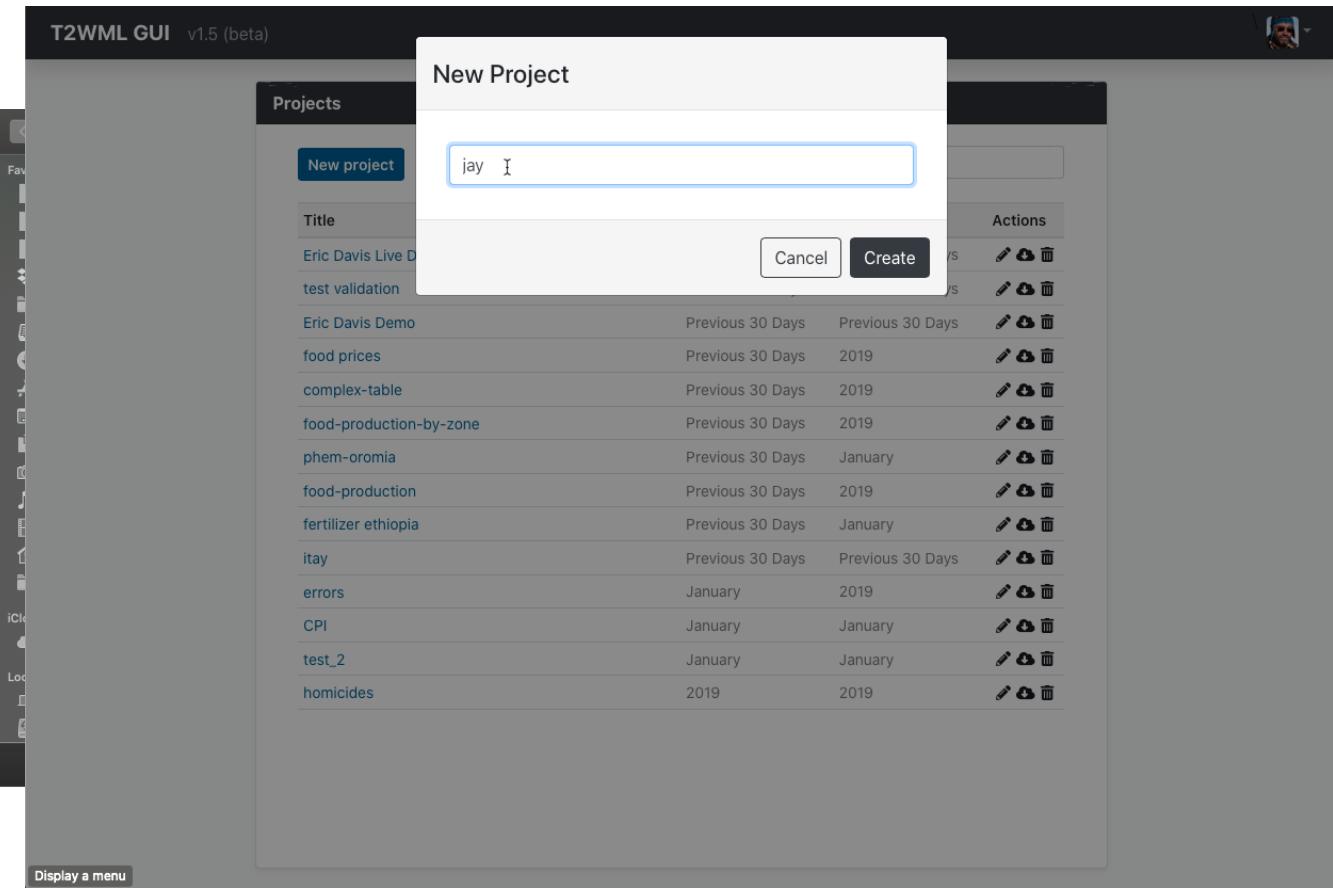
**Course Information**

Instructor Name	Number of Registered Students
Prof. Tejada	546



## F1 Macro Scores for cell role classification (In-Domain)



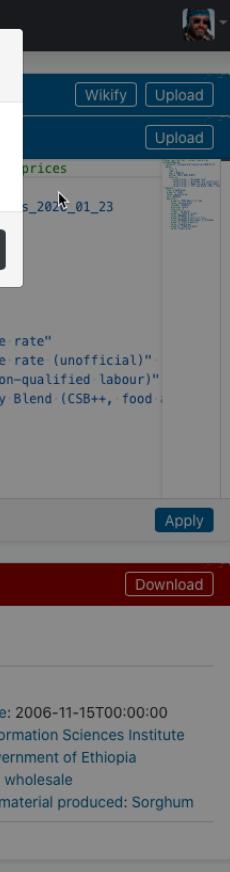


Make sure the SPARQL endpoint uses https, you can also use port 8899 if you are in the VPN



### Settings

SPARQL endpoint:



Output

N/A (Q1000000003)

price	307
- point in time:	2006-11-15T00:00:00
- curator:	Information Sciences Institute
- P7482:	Government of Ethiopia
- has quality:	wholesale
- product or material produced:	Sorghum

## T2WML GUI v1.5 (beta)



	A	B	C	D	E	F	G	H	I
	date	cname	item	name	unit	category	price	currency	country
1	7/15/05	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	238	ETB	Ethiopia
2	8/15/05	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	250	ETB	Ethiopia
3	9/15/05	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	248	ETB	Ethiopia
4	10/15/05	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	233	ETB	Ethiopia
5	11/15/05	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	252	ETB	Ethiopia
6	12/15/05	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	253	ETB	Ethiopia
7	1/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	243	ETB	Ethiopia
8	2/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	237	ETB	Ethiopia
9	3/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	245	ETB	Ethiopia
10	4/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	249	ETB	Ethiopia
11	5/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	260	ETB	Ethiopia
12	6/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	264	ETB	Ethiopia
13	7/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	275	ETB	Ethiopia
14	8/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	293	ETB	Ethiopia
15	9/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	298	ETB	Ethiopia
16	10/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	291	ETB	Ethiopia
17	11/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	307	ETB	Ethiopia
18	12/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	341	ETB	Ethiopia
19	1/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	323	ETB	Ethiopia
20	2/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	316	ETB	Ethiopia
21	3/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	318	ETB	Ethiopia
22	4/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	320	ETB	Ethiopia
23	5/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	355	ETB	Ethiopia
24	6/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	326	ETB	Ethiopia
25	7/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	338	ETB	Ethiopia
26	8/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	397	ETB	Ethiopia
27	9/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	362	ETB	Ethiopia
28	10/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	372	ETB	Ethiopia
29	11/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	378	ETB	Ethiopia
30	12/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	357	ETB	Ethiopia
31	1/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	385	ETB	Ethiopia
32	2/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	385	ETB	Ethiopia
33	3/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	430	ETB	Ethiopia
34	4/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	469	ETB	Ethiopia
35	5/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	561	ETB	Ethiopia
36	6/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	725	ETB	Ethiopia
37	7/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	740	ETB	Ethiopia

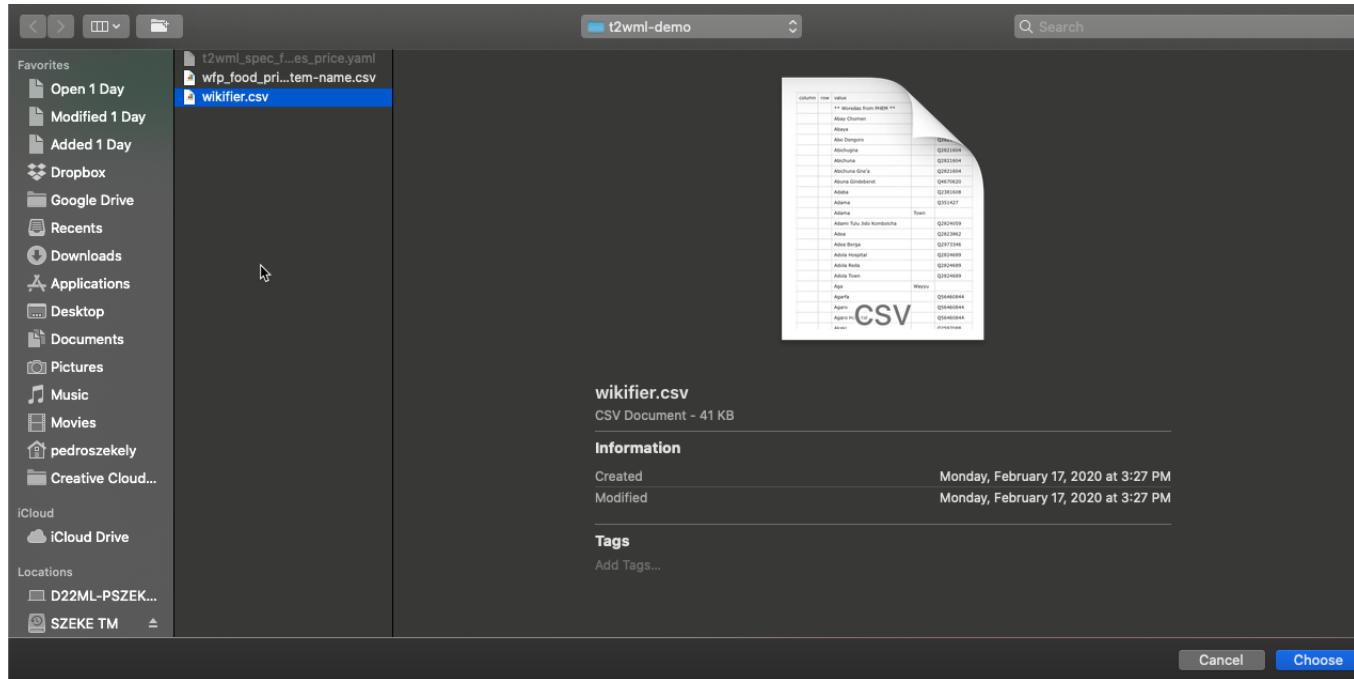
Display a menu

Wikifier			Upload	
Table	Wikidata	Wikify	Upload	
context	col 1 ↑ r... 2 ↑ value	item	label	description
No Rows To Show				

YAML Editor		Upload

Output		Download

## upload wikifier file



## T2WML GUI v1.5 (beta)



wfp\_food\_prices\_ethiopia-item-name.csv [Read-Only]

Upload

	A	B	C	D	E	F	G	H	I
1	date	cmname	item	name	unit	category	price	currency	country
2	7/15/05	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	238	ETB	Ethiopia
3	8/15/05	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	250	ETB	Ethiopia
4	9/15/05	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	248	ETB	Ethiopia
5	10/15/05	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	233	ETB	Ethiopia
6	11/15/05	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	252	ETB	Ethiopia
7	12/15/05	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	253	ETB	Ethiopia
8	1/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	243	ETB	Ethiopia
9	2/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	237	ETB	Ethiopia
10	3/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	245	ETB	Ethiopia
11	4/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	249	ETB	Ethiopia
12	5/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	260	ETB	Ethiopia
13	6/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	264	ETB	Ethiopia
14	7/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	275	ETB	Ethiopia
15	8/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	293	ETB	Ethiopia
16	9/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	298	ETB	Ethiopia
17	10/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	291	ETB	Ethiopia
18	11/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	307	ETB	Ethiopia
19	12/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	341	ETB	Ethiopia
20	1/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	323	ETB	Ethiopia
21	2/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	316	ETB	Ethiopia
22	3/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	318	ETB	Ethiopia
23	4/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	320	ETB	Ethiopia
24	5/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	355	ETB	Ethiopia
25	6/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	326	ETB	Ethiopia
26	7/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	338	ETB	Ethiopia
27	8/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	397	ETB	Ethiopia
28	9/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	362	ETB	Ethiopia
29	10/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	372	ETB	Ethiopia
30	11/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	378	ETB	Ethiopia
31	12/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	357	ETB	Ethiopia
32	1/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	385	ETB	Ethiopia
33	2/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	385	ETB	Ethiopia
34	3/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	430	ETB	Ethiopia
35	4/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	469	ETB	Ethiopia
36	5/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	561	ETB	Ethiopia
37	6/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	725	ETB	Ethiopia
38	7/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	740	ETB	Ethiopia

Display a menu

Wikifier

Wikify

Upload

YAML Editor

Upload

```

1  ### A simplest sample of T2WML.
2  ### Replace all #PLACEHOLDER below to start.
3  statementMapping:
4    region:
5      left: #CHAR
6      right: #CHAR
7      top: #INT
8      bottom: #INT
9      template:
10     item: #EXPRESSION/QNODE
11     property: #EXPRESSION/PNODE
12     value: #EXPRESSION/VALUE
13     qualifier:
14       property: #EXPRESSION/PNODE
15       value: #EXPRESSION/VALUE

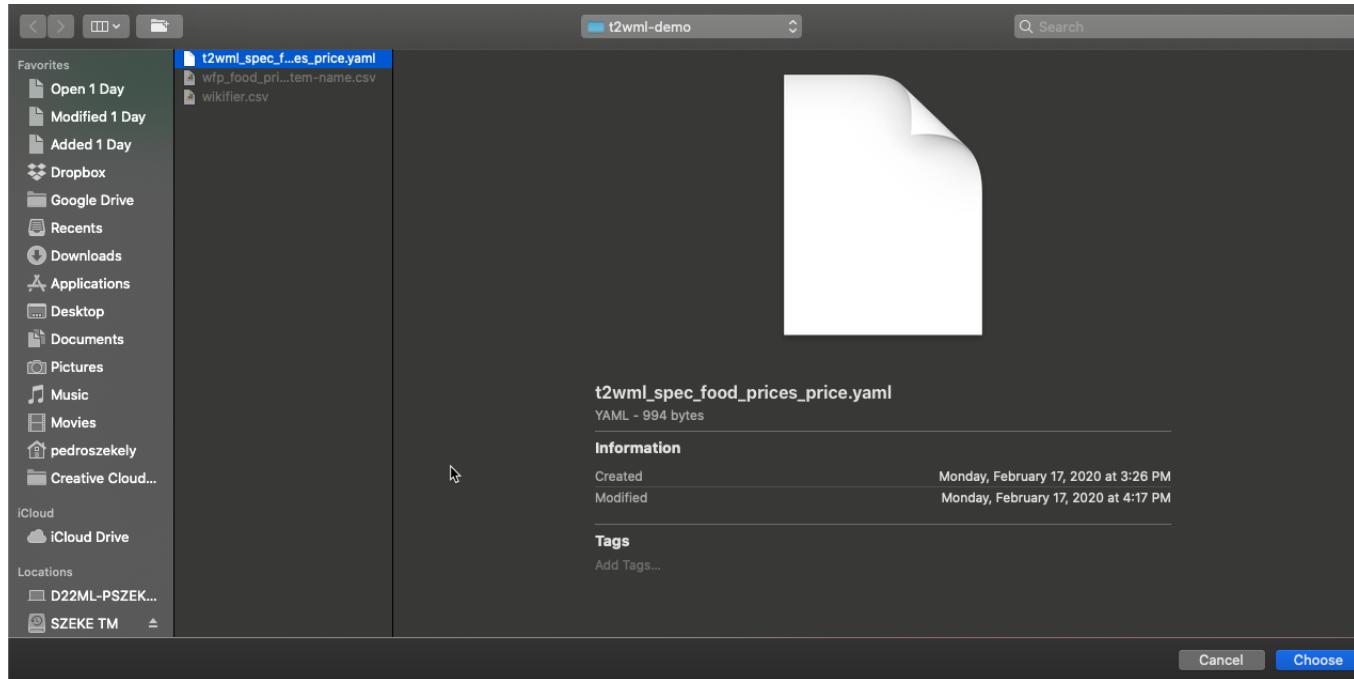
```

Apply

Output

Download

## Upload YAML



## T2WML GUI v1.5 (beta)

	A	B	C	D	E	F	G	H	I
1	date	cmmname	item	name	unit	category	price	currency	country
2	7/15/05	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	238	ETB	Ethiopia
3	8/15/05	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	250	ETB	Ethiopia
4	9/15/05	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	248	ETB	Ethiopia
5	10/15/05	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	233	ETB	Ethiopia
6	11/15/05	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	252	ETB	Ethiopia
7	12/15/05	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	253	ETB	Ethiopia
8	1/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	243	ETB	Ethiopia
9	2/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	237	ETB	Ethiopia
10	3/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	245	ETB	Ethiopia
11	4/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	249	ETB	Ethiopia
12	5/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	260	ETB	Ethiopia
13	6/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	264	ETB	Ethiopia
14	7/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	275	ETB	Ethiopia
15	8/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	293	ETB	Ethiopia
16	9/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	298	ETB	Ethiopia
17	10/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	291	ETB	Ethiopia
18	11/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	307	ETB	Ethiopia
19	12/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	341	ETB	Ethiopia
20	1/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	323	ETB	Ethiopia
21	2/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	316	ETB	Ethiopia
22	3/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	318	ETB	Ethiopia
23	4/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	320	ETB	Ethiopia
24	5/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	355	ETB	Ethiopia
25	6/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	326	ETB	Ethiopia
26	7/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	338	ETB	Ethiopia
27	8/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	397	ETB	Ethiopia
28	9/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	362	ETB	Ethiopia
29	10/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	372	ETB	Ethiopia
30	11/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	378	ETB	Ethiopia
31	12/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	357	ETB	Ethiopia
32	1/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	385	ETB	Ethiopia
33	2/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	385	ETB	Ethiopia
34	3/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	430	ETB	Ethiopia
35	4/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	469	ETB	Ethiopia
36	5/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	561	ETB	Ethiopia
37	6/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	725	ETB	Ethiopia
38	7/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	740	ETB	Ethiopia

Display a menu

Wikifier Upload

YAML Editor Upload

```

1 # Skips rows that don't contain food prices
2 statementMapping:
3   created_by: ethiopia_wfp_food_prices_2020_01_23
4   region:
5     left: F
6     right: H
7     top: 1 #602 #1
8     bottom: 10671 #604 #10671
9   skip_row:
10    - value(C$row) = "Exchange rate"
11    - value(C$row) = "Exchange rate (unofficial)"
12    - value(C$row) = "Wage (non-qualified labour)"
13    - value(C$row) = "Corn Soy Blend (CSB++, food"
14   template:
15     item: item(M$row)
16     property: P2284
17     value: value(G$row)
18     unit: Q206243

```

Output Download

## Apply YAML

**T2WML GUI v1.5 (beta)**

	A	B	C	D	E	F	G	H	I
1	date	cmname	item	name	unit	category	price	currency	country
2	7/15/05	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	238	ETB	Ethiopia
3	8/15/05	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	250	ETB	Ethiopia
4	9/15/05	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	248	ETB	Ethiopia
5	10/15/05	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	233	ETB	Ethiopia
6	11/15/05	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	252	ETB	Ethiopia
7	12/15/05	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	253	ETB	Ethiopia
8	1/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	243	ETB	Ethiopia
9	2/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	237	ETB	Ethiopia
10	3/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	245	ETB	Ethiopia
11	4/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	249	ETB	Ethiopia
12	5/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	260	ETB	Ethiopia
13	6/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	264	ETB	Ethiopia
14	7/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	275	ETB	Ethiopia
15	8/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	293	ETB	Ethiopia
16	9/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	298	ETB	Ethiopia
17	10/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	291	ETB	Ethiopia
18	11/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	307	ETB	Ethiopia
19	12/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	341	ETB	Ethiopia
20	1/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	323	ETB	Ethiopia
21	2/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	316	ETB	Ethiopia
22	3/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	318	ETB	Ethiopia
23	4/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	320	ETB	Ethiopia
24	5/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	355	ETB	Ethiopia
25	6/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	326	ETB	Ethiopia
26	7/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	338	ETB	Ethiopia
27	8/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	397	ETB	Ethiopia
28	9/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	362	ETB	Ethiopia
29	10/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	372	ETB	Ethiopia
30	11/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	378	ETB	Ethiopia
31	12/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	357	ETB	Ethiopia
32	1/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	385	ETB	Ethiopia
33	2/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	385	ETB	Ethiopia
34	3/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	430	ETB	Ethiopia
35	4/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	469	ETB	Ethiopia
36	5/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	561	ETB	Ethiopia
37	6/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	725	ETB	Ethiopia
38	7/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	740	ETB	Ethiopia

**Wikifier**

**YAML Editor**

```

1 # Skips rows that don't contain food prices
2 statementMapping:
3   - created_by: ethiopia_wfp_food_prices_2020_01_23
4   - region:
5     - left: F
6     - right: H
7     - top: 1 #602 #1
8     - bottom: 10671 #604 #10671
9   - skip_row:
10    - value(C$row) = "Exchange rate"
11    - value(C$row) = "Exchange rate (unofficial)"
12    - value(C$row) = "Wage (non-qualified labour)"
13    - value(C$row) = "Corn Soy Blend (CSB++, food"
14   - template:
15     - item: item(M$row)
16     - property: P2284
17     - value: value(G$row)
18     - unit: 0206243

```

**Output**

Click on cell

### T2WML GUI v1.5 (beta)

	A	B	C	D	E	F	G	H	I
1	date	cminame	item	name	unit	category	price	timestamp	country
2	7/15/05	Sorghum...	Sorghum	Wholesale	100 KG	{ \$col: G, \$row: 18 }		Pinned	a
3	8/15/05	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	290	ETB	Ethiopia
4	9/15/05	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	248	ETB	Ethiopia
5	10/15/05	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	233	ETB	Ethiopia
6	11/15/05	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	252	ETB	Ethiopia
7	12/15/05	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	253	ETB	Ethiopia
8	1/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	243	ETB	Ethiopia
9	2/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	237	ETB	Ethiopia
10	3/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	245	ETB	Ethiopia
11	4/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	249	ETB	Ethiopia
12	5/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	260	ETB	Ethiopia
13	6/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	264	ETB	Ethiopia
14	7/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	275	ETB	Ethiopia
15	8/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	293	ETB	Ethiopia
16	9/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	298	ETB	Ethiopia
17	10/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	291	ETB	Ethiopia
18	11/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	307	ETB	Ethiopia
19	12/15/06	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	341	ETB	Ethiopia
20	1/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	323	ETB	Ethiopia
21	2/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	316	ETB	Ethiopia
22	3/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	318	ETB	Ethiopia
23	4/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	320	ETB	Ethiopia
24	5/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	355	ETB	Ethiopia
25	6/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	326	ETB	Ethiopia
26	7/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	338	ETB	Ethiopia
27	8/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	397	ETB	Ethiopia
28	9/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	362	ETB	Ethiopia
29	10/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	372	ETB	Ethiopia
30	11/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	378	ETB	Ethiopia
31	12/15/07	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	357	ETB	Ethiopia
32	1/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	385	ETB	Ethiopia
33	2/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	385	ETB	Ethiopia
34	3/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	430	ETB	Ethiopia
35	4/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	469	ETB	Ethiopia
36	5/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	561	ETB	Ethiopia
37	6/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	725	ETB	Ethiopia
38	7/15/08	Sorghum...	Sorghum	Wholesale	100 KG	cereals a...	740	ETB	Ethiopia

### Wikifier

Upload

Wikify Upload

### YAML Editor

```

1 # Skips rows that don't contain food prices
2 statementMapping:
3   created_by: ethiopia_wfp_food_prices_2020_01_23
4   region:
5     left: F
6     right: H
7     top: 1 #602 #1
8     bottom: 10671 #604 #10671
9   skip_row:
10    - value(C$row) = "Exchange rate"
11    - value(C$row) = "Exchange rate (unofficial)"
12    - value(C$row) = "Wage (non-qualified labour)"
13    - value(C$row) = "Corn Soy Blend (CSB++, food"
14   template:
15     item: item(M$row)
16     property: P2284
17     value: value(G$row)
18     unit: Q10671

```

Apply

### Output

Download

N/A (Q1000000003)

price	307
- point in time: 2006-11-15T00:00:00	
- curator: Information Sciences Institute	
- P7482: Government of Ethiopia	
- has quality: wholesale	
- product or material produced: Sorghum	