

Atomizing the Chicago Assyrian Dictionary: Agentic Workflow for Structuring Multi-Language Text Documents

Presenter: Himi Yadav
Research Data Scientist, Research Computing Center

In collaboration with:
Hervé Reculeau and Foy Scalf
Institute of the Study of Ancient Cultures

Tuesday, September 15, 2026



eCAD Team

ISAC

RCC

Stauffer



Hervé Reculeau



Foy Scalf



Me



Mohsen Zand



Afraaz Ali



Christian Borgen



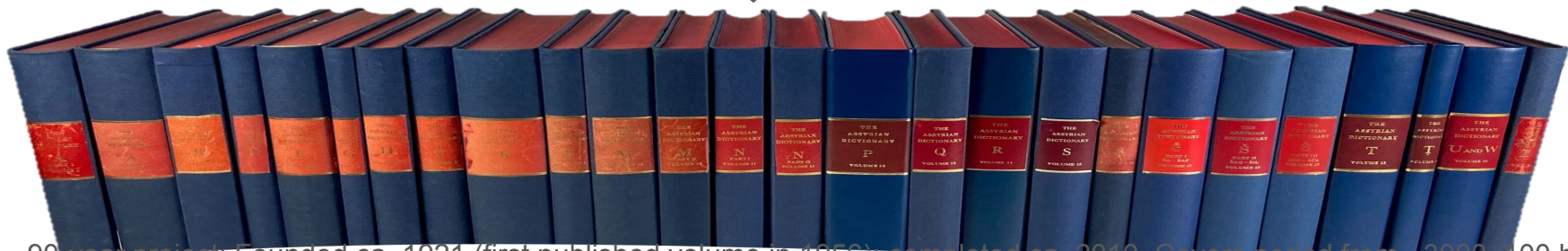
Anthony Ray



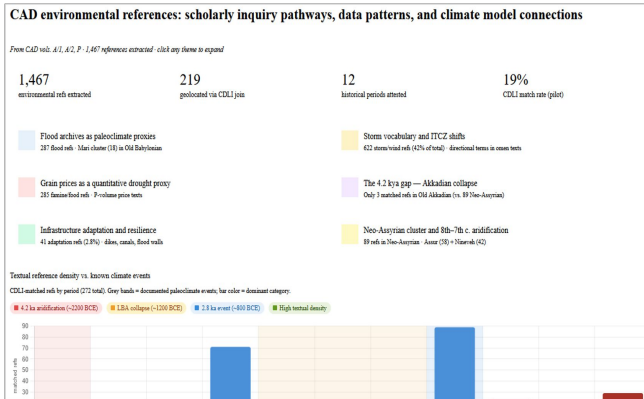
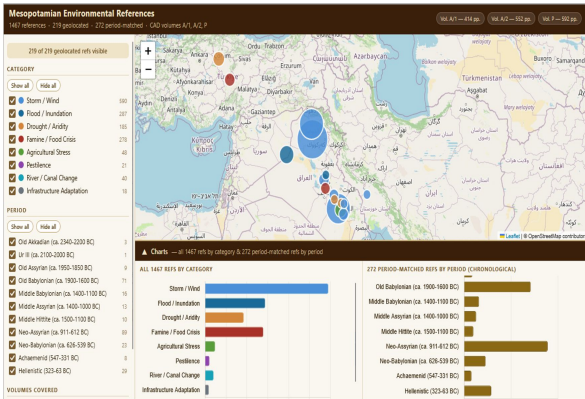
Daniel Blandes



Scott Bell



90-year project. Founded ca. 1921 (first published volume in 1956), completed ca. 2010, Covers period from ~3000–100 bce



Corpus Creation Task at Low-Medium Compute Scale

The screenshot shows the homepage of the Chicago Assyrian Dictionary. At the top, there is a navigation bar with links: DICTIONARY, BROWSE, ARCHIVES, RESEARCH, ABOUT, and FAQ. Below this, the header identifies the project as a collaboration between the University of Chicago and the Institute for the Study of Ancient Cultures. The main title, "CHICAGO ASSYRIAN DICTIONARY", is prominently displayed, followed by the tagline "Three millennia of Akkadian, modeled for the digital age". A search bar is located in the center, featuring a "LEMMA" dropdown and a "SEARCH" button. Below the search bar, a "POPULAR" section lists several lemmas: *lebu*, *šamu*, *nāru*, *bīru*, *māru*, *ilu*, and *šamū*. A "CORPUS STATS" section is presented with four large, dark boxes containing the following data: 26 VOLUMES, 28,000+ LEMMA, 3,000 yrs ATTESTED SPAN, and 300K+ ATTESTATIONS. Below this, a "BROWSE BY VOLUME" section features a horizontal row of buttons for each letter of the alphabet (A through Z), with additional buttons for I/J, U/W, and a final Z button. The footer contains the logos of the University of Chicago and the Institute for the Study of Ancient Cultures, along with the project's name and a set of links: Link 1, Link 2, Link 3, and Link 4.

PREPOSITION

adi A

5 senses | 34 usages | Volume A1

MEANINGS & ATTESTATIONS

The meanings attested for this lemma, with translations and citations.

1. *up to, as far as*

WITH GEOGRAPHICAL NAMES

a passim; note with specifications: GN *a-di abullīši* GN as far as its city gate EA 106:11, *a-di-i mišir* URU *Girsu* CT 36 7 ii 1, and dupl. BIN 2 33, *a-di qereb Sippar* VAB 4 166 B vi 68 (Nbk.), *a-di tāmert* Ninua OIP 2 98:89 (Senn.); *a-ti-ma pūti tiāmtim* Hirsch, AfO 20 37 vi 8 (Sar.); note with ethnic names: *a-di LÚ Litamu italak* he went as far as (the region of) the Litamu tribe ABL 436:9.

WITH TOPOGRAPHIC INDICATIONS

b *a-di reš ID e-ni ša Idiglar* as far as the source of the Tigris Layard 90:69 (Shalm. III), cf. *a-di bāb(!) ID marra* ABL 418 r. 5, *a-di-i mušannitu ša* GN as far as the dike of GN BE 9 59:3 and 13 (NB), *a-ti abul nišim* Belleten 14

Corpus Creation

Lemmatized Online Corpora

Assyrian:

 History Department
Old Assyrian Research Environment

Text Corpus of Middle Assyrian

Archival Texts of the Assyrian Empire (ATAE)

State Archives of Assyria Online

Specialized corpora:

The Royal Inscriptions of Babylonia online (RIBo) Project

The Royal Inscriptions of Assyria online (RIAo) Project

The Royal Inscriptions of the Neo-Assyrian Period

Oracc: Glass — Corpus of Glass Technological Texts



Cuneiform Digital
Library Initiative

Babylonian:

ARCHIVES BABYLONIENNES (XX^e-XVII^e SIÈCLES AV. J.-C.)

ARCHIBAB



Info ▾ DigEanna ▾ Navico ▾ Nippur ▾ Browse ▾ API

ACHEMENET

L'empire perse achéménide
du Scythie à l'Indus
de 550 à 330 avant notre ère



mon espace

Hellenistic Babylonia: Texts, Images, Names

electronic
babylonian
Library



Efforts so far ... on mostly CPUs



DASHBOARD

CAD_A_1.pdf

Page 151

READ

EDIT

HISTORY

SAVE

COMPLETE

<

>

C... 151 / 428 - 97% +

oi.chicago.edu

adi A

informing my lord YOS 3 45:8, cf. *a-di-i la-i uttatu tamattú* YOS 3 81:26 and 45:21, cf. *a-di la-i* BRM 2 47:23 and ZA 2 173:9 (all NB).

c) in the NB oath formula to introduce a promissory oath: PN ina Bēl Nabū Bēlti-ša-Uruk u Nanā adē ša RN u PN₂ mār šarri ittemi kī a-di-i UD.7.KAM ... ana Uruk ... allakamma PN took an oath by Bēl, Nabū, the Lady-of-Uruk and Nanā (and) by the majesty(?) of Nabonidus and Prince Belshazzar: I shall come to Uruk on the seventh (of MN) PSBA 38 27:4, and passim, cf. PN ina ... *a-di-e-ša RN ana PN₂ ittemi kī a-di-i* VAS 6 118:4, TCL 13 163:16, also TuM 2-3 120:10, JRAS 1926 107:10, Nbn. 197:7, and passim; ^dEN u ^{d(1)}AG(!) kī-i a-di ABL 1236:21 (NB); note the variations: *kī-i a-di-ia* Dar. 309:3, Dar. 339:5 and BOR 1 102:4, *a-kī-i a-di-i* AJSL 27 224 No. 21:4, without kī ZA 4 151 No. 7:5, VAS 4 13:7; note *šumu ili ana aḫāmeš ultēlu umma kī a-di fēmu* ... nišapparakka they have taken a mutual oath as follows: we shall report to you (whatever we hear) ABL 282 r. 5.

adi A (*adu, gadu, qadu, ḥadu, ad*) prep.; 1. up to, as far as (connoting space or distance), 2. until, up to (connoting time), 3. concerning, as to, 4. (with numerals) times, -fold, for (the nth) time, into (n) parts, 5. in adverbial expressions; from OAkk., OA and OB on; *adum* in OAkk., OA, *qadu* AOB 1 34:8 and 10, MRS 9 229:15', *adu* passim in RS, Bogh., MA, NA, *gadu* BE 39099, NB, Gössmann Era V 59, and passim in Nbk., *ḥadu* BIN 7

adi A 1b

236; u, za1.šō: *a-di ur-ra-am*, mi, za1.šō: *a-di ka-qa-tim* OBGT I 804f. zag.ki.a: *a-di al-raš* RA 12 74:31f.; a.rā 1 a.rā 2: *a-di 1-tu a-di 2-du* Ai. IV iv 28; 7 a.rā 7: 7 a-di 7 KAR 34:2 and 4; a.rā 2.kam: *a-di šina* CT 17 26:53, and passim, and note a.rā 2.kam. ma.šō: *a-di šina* Surpu V-VI 23f.; 7.na a.rā 2.[km]: *sibit a-di šina* CT 17 24 iv 232f., and passim. me.na.šō: *a-di ma-ti* OBGT I 734, me.na.šō.km: *a-di ma-ti-ma*, also, with Sum.en.na.me.šō(Akk.), *en.šō(Akk.)* ibid. 735ff.; me.en.na: *a-di ma-ti* 4R 29** 7f., me.na: *a-di ma-ti* BA 10/1 76 No. 4 r. 30f., me.na.na: *a-di ma-ti* 4R 10 r. 21f., en.me.en: *a-di ma-ti* BA 5 633 No. 6:29f., me.en.šō: *a-di ma-ti* SBH p. 64:9f., en.šō: *a-di ma-ti* SBH p. 53:23f., en.šō: *a-di ma-ti* VAS 10 179:1f.

1. up to, as far as — a) with geographical names: passim; note with specifications: GN *a-di abullūš* GN as far as its city gate EA 109:11, *a-di-i mišir Uku Girsu* CT 36:7 ii 1, and dupl. BIN 2 32, *a-di qereš Šippar* VAB 4 166 B vi 68 (Nbk.), *a-di tāmerti Ninna* OIP 2 98:89 (Sonn.); *a-ti-ma pūti tiāntim* Hirsch, AFO 20 37 vi 8 (Sar.); note with ethnic names: *a-di lū Litamu italak* he went as far as (the region of) the Litamu tribe ABL 436:9.

b) with topographic indications: *a-di rēš* in *e-ni ša Diglat* as far as the source of the Tigris Layard 90:69 (Shalm. III), cf. *a-di bāb(!)* in *marra* ABL 418 r. 5, *a-di-i mušannitu ša GN* as far as the dike of GN BE 9 59:3 and 13 (NB), *a-di ubul nišim* Beleten 14 174:9 (Irkam), *a-di kiṇai qa Puratte* AKA 354:ii 29 (Am.), *a-di titurri mētiṭti šarri* as far as the causeway of the king's passage MDP 6 pl. 11 i 3; *a-di GN* *mānādā ša kīni* as far as GN at the outlet

Last Edit: Charlotte Gilmore

5/7/2026, 11:04 AM

adi A adi A 1b

informing my lord YOS 3 45:8, cf. *a-di-i la-i uttatu tamattú* YOS 3 81:26 and 45:21, cf. *a-di la-i* BRM 2 47:23 and ZA 2 173:9 (all NB).

c) in the NB oath formula to introduce a promissory oath: PN ina Bēl Nabū Bēlti-ša-Uruk u Nanā adē ša RN u PN₂ mār šarri ittemi kī a-di-i UD.7.KAM ... ana Uruk ... allakamma PN took an oath by Bēl, Nabū, the Lady-of-Uruk and Nanā (and) by the majesty(?) of Nabonidus and Prince Belshazzar: I shall come to Uruk on the seventh (of MN) PSBA 38 27:4, and passim, cf. PN ina ... *a-di-e-ša RN ana PN₂ ittemi kī a-di-i* VAS 6 118:4, TCL 13 163:16, also TuM 2-3 120:10, JRAS 1926 107:10, Nbn. 197:7, and passim; ^dEN u ^{d(1)}AG(!) kī-i a-di ABL 1236:21 (NB); note the variations: *kī-i a-di-ia* Dar. 309:3, Dar. 339:5 and BOR 1 102:4, *a-kī-i a-di-i* AJSL 27 224 No. 21:4, without kī ZA 4 151 No. 7:5, VAS 4 13:7; note *šumu ili ana aḫāmeš ultēlu umma kī a-di fēmu* ... nišapparakka they have taken a mutual oath as follows: we shall report to you (whatever we hear) ABL 282 r. 5.

adi A (*adu, gadu, qadu, ḥadu, ad*) prep.; 1. up to, as far as (connoting space or distance), 2. until, up to (connoting time), 3. concerning, as to, 4. (with numerals) times, -fold, for (the nth) time, into (n) parts, 5. in adverbial expressions; from OAkk., OA and OB on; *adum* in OAkk., OA, *qadu* AOB 1 34:8 and 10, MRS 9 229:15', *adu* passim in RS, Bogh., MA, NA, *gadu* BE 39099, NB, Gössmann Era V 59, and passim in Nbk., *ḥadu* BIN 7 90:13, *ad kimaši ūmē* Tn.-Epic "ii" 15, "iii" 13, for *ad* with *mati* see mng. 2g-2'; wr. syll. (*a-di*/*di-i* passim in OA, NB) and EN (in mng. 4 also A.RÁ); cf. *adi* conj., *adimi*, *adišu*.

en ZAG = *a-di* A VIII/3:35; en = *a-di* Igituh short version 26; en.na = *a-di* AN.TA NBGT II 41; za-ag ZAG = *a-di* Idu 1 158, also S⁹ Voc. AE 15'; zag = *a-di* CT 18 50 r. ii 3; x.za = *a-di* = (Hitt.) ku-it-ma-an Erimhuš Bogh. A 17; [za-ag] [ZAG] = [a]-[du]-u A VIII/4:5; [A.DU] = *a-*

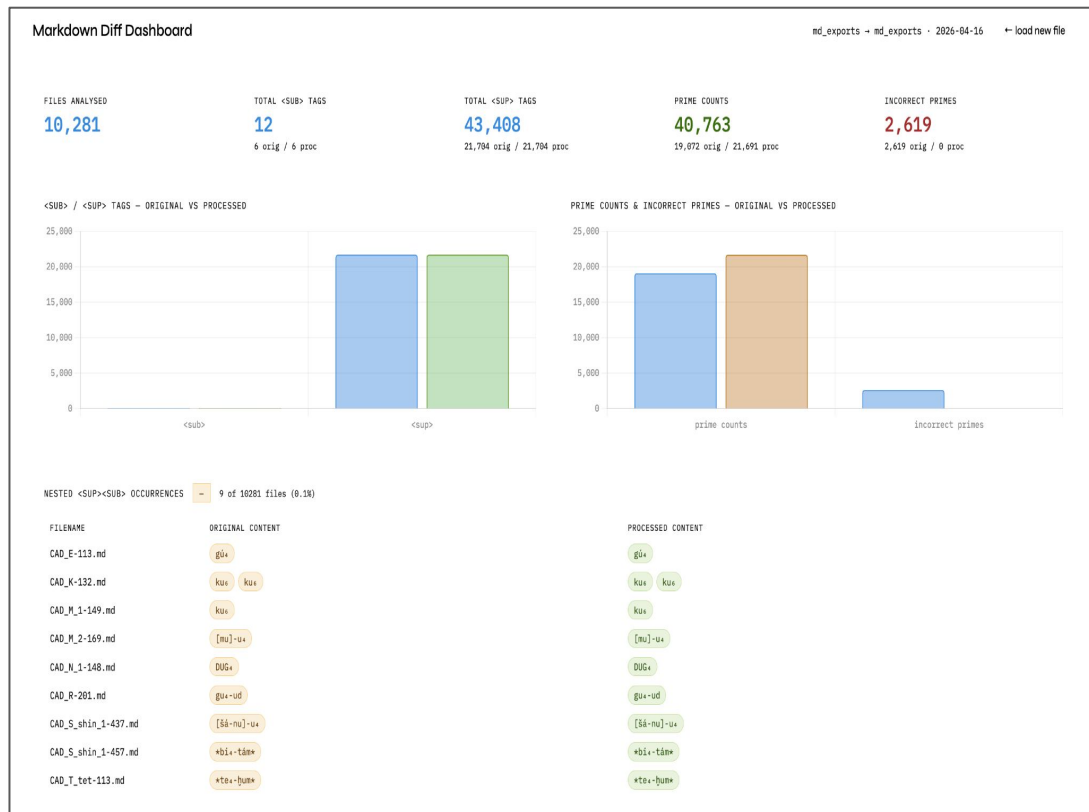
Version History

- C Charlotte Gilmore 5/7/2026, 11:04 AM
- Original Import 3/25/2026, 9:01 AM

Version History

- H Herve Reculeau 7/8/2026, 7:16 AM
- P Post Processor Script 4/17/2026, 1:19 PM
- C Charlotte Gilmore 4/5/2026, 10:39 AM
- C Charlotte Gilmore 4/2/2026, 11:23 AM
- Original Import 3/25/2026, 9:01 AM

Lessons learnt from image to text task (Proprietary VLMs/HITL)



- VLM failures on **niche Akkadian transliteration standards** vs common proprietary model development data peculiarities
- Visibility and traceability of VLM errors
- Prompt brittleness and Model Performance vs Human-Expert Ground Truth Agreement (Not many annotator disagreements yet)
- Indirect challenges
 - Compute and developer cost of open-source model development vs breadth of AI-skilled users for niche open-source models vs data users
 - Proprietary model development transparency

Task at Hand: Atomize Dictionary

- Unstructured text to structured text
- CAD structure is specific to the CAD
- Output: Relational Database (PostgreSQL)

Zero-shot Prompting

ṭābtu A ṭābtu A

(U+TIM) ibid. 203 M.12109-3, ibid. 205 M.18071+7, 15, and 17.

3^r in MB, MA: preparation should be made for the feast *ṭābtu* (gizlū) MUN u mērestu mala ibasū lu māda there should be plenty of wood, twigs(?) torches, salt, and whatever else is desirable Aro, VZJ 8 567 HS 110-11 (MB let.); MUN (column heading) PBS 2/2 91-1; 2 (BĀN) MUN (between flour and barley) BE 15 181-9, cf. BE 15 44-20, 22, and 35 (all MB adm.); MUN *laṣu* there is no salt available VAS 19 71-25 (MA let.).

4^r in Nuzi, RS: 3 ANŠE ŠE.BAR ana *ṭim* ana 3 ANŠE *ṭā-ab-i* three homers of barley for the price of three homers of salt HSS 13 44-11, cf. ibid. 42-4, HSS 15 255-5; uncert.: *bani akanna ana pa-ri* ¹ka *ki jānu karānu u* GĪŠ.I.GĪŠ.MEŠ ḪA.LA.MEŠ u MUN.MEŠ does it seem proper to you that there is no wine, or oil(?), or *ṭālu* vinegar, or salt? Ugaritica 5 54-17.

5^r in NA, NB: 2 GUN MUN ana 2 GĪN KÜ.BABBAR Nbk. 447-1; x KÜ.BABBAR ana MUN.HI.A ana PN nadin CT 57 162-8, cf. CT 56 153-7 (both NB); 1-2 SĪLA I.MEŠ 1 (BĀN) kudšimne 1 (BĀN) MUN.MEŠ (among supplies for charioteers) Iraq 15 146 ND 3467-9, see Postgate Taxation 399; 8 SĪLA 1/2 qār MUN ŠE kudšimne atidin ABL 207 r. 2, see Parpola, SAA 5 242, cf. CT 53 27-4; see Fuchs and Parpola, SAA 15 74 (all NA); 52 MA.NA MUN PN ... *maḫi* (f) Nbk. 155-1; 15 MA.NA MUN.HI.A PN AnOr 8 33-1, cf. CT 56 783-11; 1 GUN 20 MA.NA MUN.HI.A ... ana PN ... nadin CT 56 604-2; note measured by volume: x GUR 2 (BĀN) 3 SĪLA MUN.HI.A Nbn. 1017-7, cf. UCP 9 91 No. 24-28; 4 GĪN ana MUN.HI.A four shekels (of silver) for salt CT 4 144-4, cf. BRM 1 99-17; *biṭṭa* ana MUN.HI.A *saḫē* an eighth of a shekel (of silver) for salt and cress seed Nbn. 173-6; [x *saḫē* ana MUN.HI.A nadin CT 56 773-43; x (BĀN) ŠE.BAR MUN.HI CT 55 62-1; MUN *ki-ba-tum* Camb. 295-2; unclear: 5 MA.NA MUN.HI.A ana *ṭi-i* *pi* CT 56 607-6; 1 PI *ṣamni* 2 PI MUN.HI.A 2 PI *saḫē* *ṣidtu* *ša* 3.TA *ṣanāti* one *pānu* of oil, two *pānu* of salt, and two *pānu* of cress seed (among other items), provisions for three years Dar. 253-8, cf. TCL 9 85-20; VAS 6 202-5, CT 56 650-9; PN ... *umu* 4 SĪLA *akalu* 3 SĪLA *ḫanū* *ina* *ṣatti* 1 PI *ṣammasammi* 1 PI MUN.HI.A x *saḫē* ana ¹PN *aṣṣatiku* u PN *māriku* inandin PN will give to his wife PN, and her son PN, a daily allowance of four shekels of bread and three shekels of beer, and a yearly allowance of 15 minas of wool, one *pānu* of linseed, one *pānu* of salt, and x cress seed Nbn. 113-4; *ṣamnu* MUN.HI.A *saḫē* u *ṣipati* *ša* *pābe* *aṣ* *qēmē* *ša* 2-ta *ṣanāti* *nikkasū* *epuṣma* ... *bēli* *ṣuḫbi* *la* *make* an accounting of the oil, salt, cress seed, and wool for those workmen, and of the two years' worth of flour, so that my lord can send them YOS 3 31-9, cf. CT 22 2-16 (all NB); for other refs., see *saḫē* mng. 2b-1; note in symbolic use, expressing reconciliation: may the king protect *amēṭūti* *ša* MUN.HI.A *ša* *mār* *Jakini* *ṭipnu* all those who have partaken of the salt of the tribe of Jakin (and spoken of peace) ABL 747 r. 6 (NB).

b) in lit., hist., inc., omens: *kima* MUN *zēra* *la* *iṣū* ... *kima* MUN *zēra* *la* *niṣū* just as salt has no seed, so, like salt, may we have no offspring KBo 1 3 r. 32 and 34 (treaty); *ṣumma* MUN *ina* *namē* *all* [*ṣanamu*] if salt is seen in the outlying area of a town CT 40 45e-10; *ṣumma* *ina* *ḫitil* *all* MUN [*ṣanamu*] CT 39 21-159 (both Sili Anu); *ina* *all* MUN *ṣammar* CT 29 48-20; *ṣumma* *ṭind* *ṭarṭi* *kima* *idru* MUN [*ṣammatu*] (see *idru* A usage ii) CT 39 14-17; *iṣṣāṣa* *lu* MUN ... *ša* *ṣuḫū* *kiṣpi* *kima* (var. *adda* *kiṣpi*) MUN (var. *adda* *ina* *mē*) *iṣṣamit* may her tongue be salt, may she who put a spell on me dissolve like (a lump of) salt (in water) Maqlu 1 31 and 33, var. from STT 78-33; *ṣumma* *ina* MN Adad *niḫmū* *iṣṣima* *erṣetu* MUN *lu* [*ṣamnu*] *lu* *iṣṣa* *lu* *kupra* *iṣṣa* if there is a thunderclap in Ajaru and the earth exudes salt or oil or crude bitumen or bitumen

12

Code

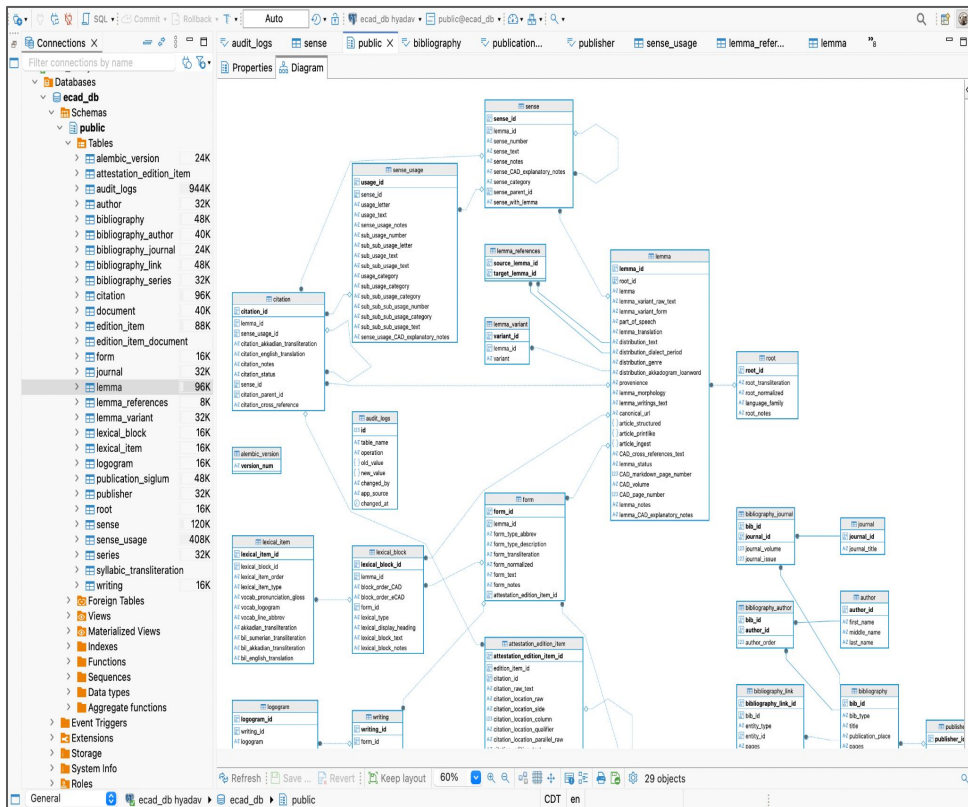
Blame

Raw



```
1 Extract all dictionary entries from this page image as JSON. Return ONLY valid JSON with no markdown or preamble.
2
3 Structure (use null for missing fields):
4 [
5   {
6     "Lemma": "headword",
7     "PartOfSpeech": "noun/verb/adj/etc",
8     "Homonym": "I/II/III or null",
9     "Variants": ["variant1", "variant2"],
10    "Distribution": "OB/MB/NA/NB/etc",
11    "Provenience": "geographic origin",
12    "Akkadogram": true/false,
13    "Morphology": "morphological analysis",
14    "Writings": ["cuneiform spelling 1", "spelling 2"],
15    "CrossReferences": ["see X", "cf. Y"],
16    "Lexical": "lexical list mentions",
17    "MeaningNumber": 1,
18    "Translation": "English gloss",
19    "UsageLetter": "a/b/c or null",
20    "Usage": "specific usage context",
21    "Quotation": "Akkadian text (transliteration)",
22    "QuoteTranslation": "English translation",
23    "QuoteCitation": "source citation",
24    "Bibliography": ["ref1", "ref2"],
25    "Discussion": "philological notes"
26  }
```

Human-Expert Structured Multi-Language Texts



PropertiesDataDiagram

Show SQL

Enter a SQL expression to filter results (Use Ctrl+Space)

Grid

Text

	Az lemma	Az CAD_volume	Az lemma_variant_raw_text	Az part_of_speech	Az lemma_translation	Az distribution_text
5	**ammaki**	A 2	*ammaku*	conjunction	instead of	SB)
6	**amru**	A 2	[NULL]	adjective	**1** checked, accounted for, **2** selected	OB, SB, NA, NB
7	**a**	A 1	**a**	demonstrative pronoun	that, those, the afore-mentioned	NB
8	**aba**	A 1	*abu*	interjection	what	SB
9	**qabū A**	Q	qabū'u*	noun	**1** speech, statement, **2** address (to a god), **3 from OB on	
10	**qabku**	A 1	[NULL]	adjective	prisoner, captive	SB, NB
11	**adi**	A 1	*adu, qadi, qadu*	conjunction	as long as, while, until, (with negation) before	from OAkk. on
12	**adi A**	A 1	*adu, qadu, qadu, ad*	preposition	**1** up to, as far as (connoting space or distance), ** from OAkk., OA and OB on	
13	**amuru**	A 2	*amuritu*	adjective	Amorite (i.e., pertaining to the Amurru-people)	from OAkk. on
14	**adu**	A 1	*adi*	adverb	here then, now then	EA, Bogh., NB
15	**adu B**	A 1	[NULL]	s. pl. tantum, noun	majesty(?), power(?)	NA, NB
16	**adu D**	A 1	[NULL]	noun	leader	[NULL]
17	**adu A** in **bēl adē**	A 1	[NULL]	noun	servant, vassal (bound by an "adu"-agreement)	NA, SB, NB
18	**qabū B**	Q	[NULL]	noun	(a container)	OB Elam, NB
19	**ababū**	A 1	*ešabū*	noun	**1** (an administrative temple official), **2** tax due	OB
20	**qabū**	Q	*qabū*	verb	**1** to say, tell, speak, to report, to recite, to pronounce from OAkk. on, Akkadogram in Hi	
21	**qabū**	Q	[NULL]	[NULL]	[NULL]	[NULL]
22	**qabū**	Q	or *qāpu*	verb (?)	(mng. unk.)	[NULL]
23	**qāpu B**	Q	*qabū*	verb	**1** to buckle, to cave in, **2** to cause to collapse	OB, SB
24	*qāpu C**	Q	or *kāpu*	verb	to intend harm	OB*
25	**qāpu**	Q	[NULL]	[NULL]	[NULL]	[NULL]
26	*qāpu*	Q	[NULL]	[NULL]	[NULL]	[NULL]
27	**adi B**	A 1	*adu*, *qadu*, *ekadu*	preposition	together with, inclusive of, pertaining to	from OAkk. on
28	**abāzu**	A 1	[NULL]	verb	**1** to seize, to hold a person, **2** to take a wife, tc from OAkk. on	
29	*qabutu**	Q	*qabū*	noun	corral, fold	NA, NB
30	**qāpu A**	Q	*qāpu*	verb	**1** to believe, to give credence to, to trust, (in the st. from OAkk. on	

References X Value

sense (sense_lemma_id,fkey)

Grid

Text

	sense_id	lemma_id	Az sense_number	Az sense_text	Az sense_notes	Az sense_CAD_explanatory_notes
1	cbd3b755-cc13-420a-838b-7d9b4d96fa1d	9979d9ce-0ff8-4c15-b328-671a838ba33a	**1**	up to, as far as	[NULL]	[NULL]
2	bac8f683-df6d-4335-ad99-164887c70194	9979d9ce-0ff8-4c15-b328-671a838ba33a	**2**	until, up to	[NULL]	[NULL]
3	2021ede1-a2bf-4f34-ab01-965ba585017e	9979d9ce-0ff8-4c15-b328-671a838ba33a	**3**	concerning, as to (mo)	[NULL]	[NULL]
4	1fcb4b5-9174-40d7-8d5d-1b3c7909c328	9979d9ce-0ff8-4c15-b328-671a838ba33a	**4**	(with numerals) times	[NULL]	[NULL]
5	3ed5affe-7fe6-4b93-ac79-4bf2a7e0a1f4	9979d9ce-0ff8-4c15-b328-671a838ba33a	**5**	*a-ti da-ni-iš* greatly	[NULL]	[NULL]

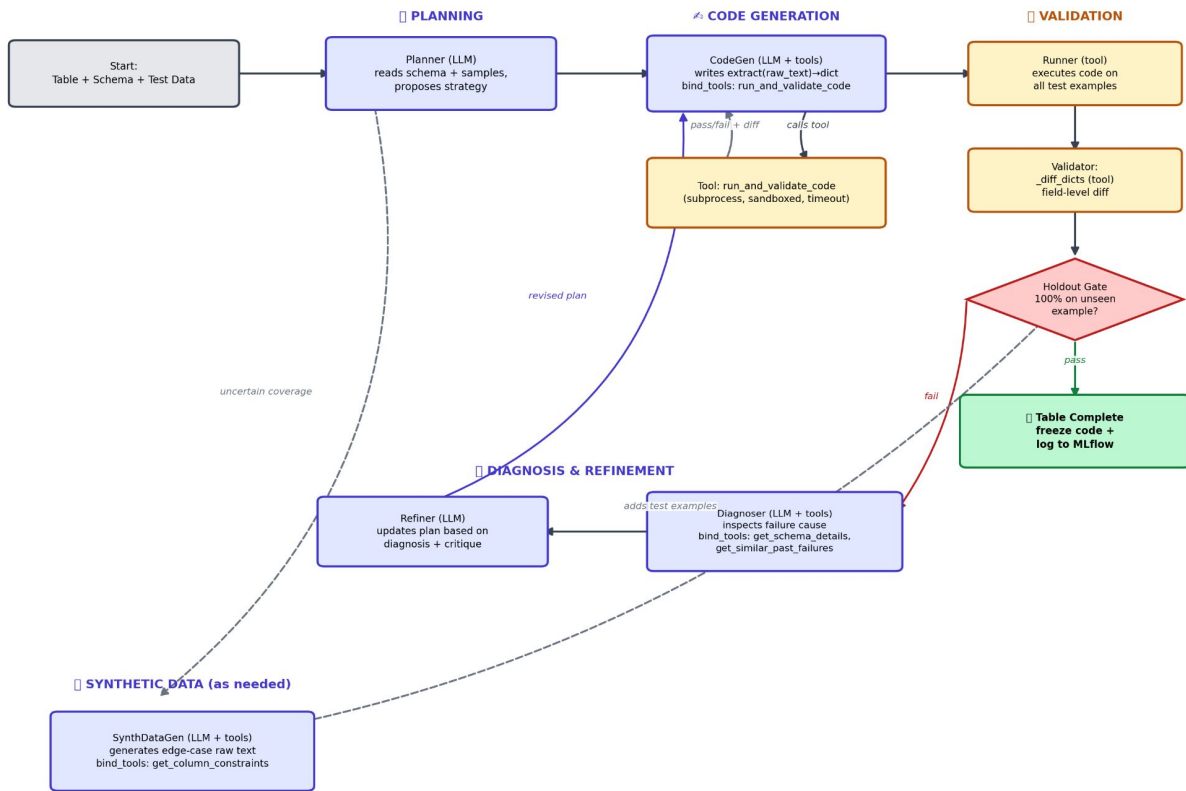
Human-Expert Manual Atomization Efforts

- DBeaver server on Midway3 CPU compute nodes (cVPN access)
- Human in the loop
 - With schema changes ~twice a week at this point
- Challenges
 - Time-consuming by hand
 - Structure is inherently present in the dictionary
 - Structure may be inconsistent, given the volumes were published from 1950s to early 2000s, with editor/data curator changes

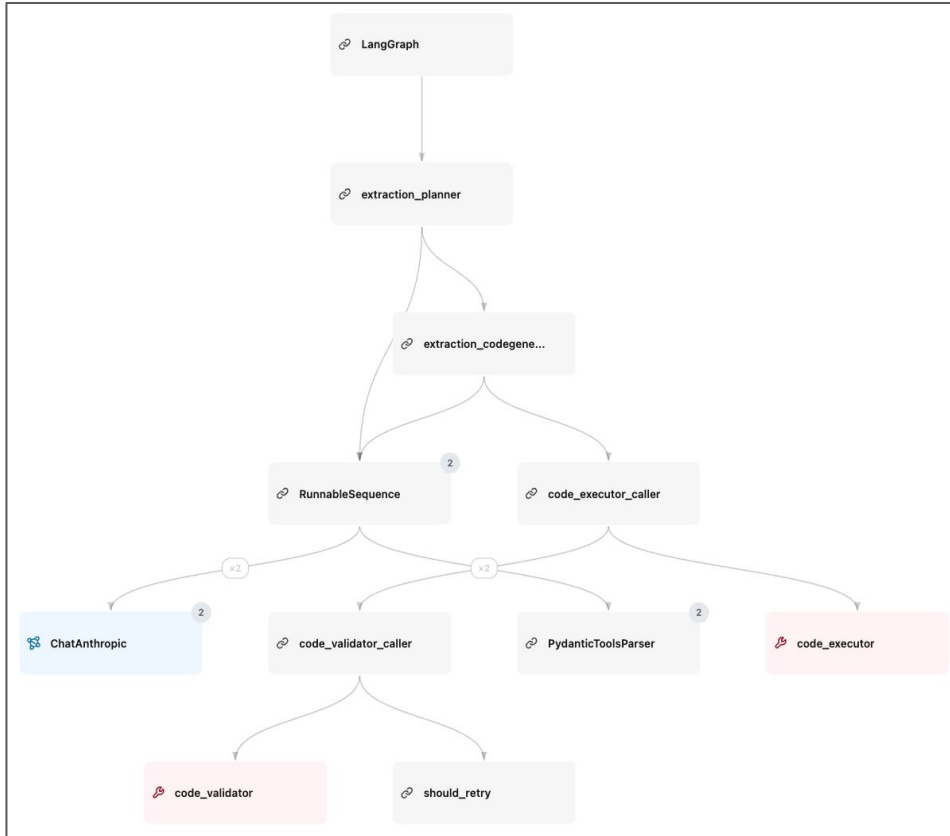
Agentic Atomization Per Table

- Problem: Given Unstructured Markdown Text Files, constrained by Table Pydantic Model
- Output
 - Generated Code
 - List Expected Outputs (constrained to table fields) per page
- Why
 - Faster
 - Generate test cases/synthetic data

Text Extraction Agent Pipeline



Academy Agent/Actors with Parsl Local Executor



```
class CodeGenState:
    iterations: int = 0
    seed_text: str = ''
    plan: str = ''
    expected_output: list[dict] = field(default_factory=list)
    generated_code: Optional[Code] = None
    code_valid_check: bool = False
    code_error: str = ''
    state_errors: str = ''
    validation_score: float = 0.0
    critique: str = ''
```

- Prototype
 - Number of agents: 2 (for Parsl)
 - Seed: 1 input example text file
 - Execution time per CodeGenAgent agent: 3-6 minutes
 - Max Iterations needed so far: 3
 - API cost per agent per table: <2\$
 - Running on
 - local CPU server (not Midway2/3 yet)

mlflow3.15.1

GenAIModel training

atomizer-extracti...

Overview

Observability

Traces

Sessions

Evaluation

Judges

Review

Datasets

Evaluation runs

Prompts & versions

Playground

Prompts

Agent versions

AI GatewayNew

Docs

Settings

Traces - E

Run 14c24

Run c26a4

Run c26a4

Run 4b93

Run ba6d

Run 0352

Run 27de

Run 27de

+

localhost:5003/#/experiments/1/traces?startTimeLabel=CUSTOM&startTime=2026-08-22T04%3A16%3A07.083Z&endTime=2026-0...

Finish update

Ask Gemini

Experiments

Traces

Custom

Search

Trace breakdown

Graph

Filter

Inputs / Outputs

Attributes

Events

Show assessments

Summary

Details & Timeline

Linked prompts

Search

Trace breakdown

LangGraph

extraction_planner

RunnableSequence

ChatAnthropic

PydanticToolsParser

extraction_codegenenerator

RunnableSequence

ChatAnthropic

PydanticToolsParser

code_executor_caller

code_executor

code_validator_caller

code_validator

should_retry

Click a graph node to navigate spans

Iterations: 0, seed_text: "# amašša amat ek...

Analyze with Assistant

Add to dataset

Flag for review

Share

Status OK

ID tr-a00f5ad59b9a875a12f829d001462a5e

Token count 32258

Cost \$0.51013

Latency 3.45m

New! Flag traces for review and add them to a review queue.

Inputs

iterations

0

seed_text

amašša amat ekalli

foreign wrd; wr. syll. and NA₄AMAŠ.PA.Ē and NA₄AMAŠ.MA₄.A.

na₄amaš.pa.è.a Nippur Forerunner to Hh. XVI 71; na₄GÚ.pa.è.a, na₄GÚ.bi.a Late OB Foreruner 72 f.

NA₄AMAŠ.PA.Ē (among stones for magic use) CT 23 37 K.2354 + iv 10, also (in an enumeration of stones) Lugale XII 21 (Sum. only, = Bergmann Lugale 514), Biggs Šaziga 67 iii 51, Köcher BAM 316 ii 2; NA₄AMAŠ.MA₄.A KA...

See more

plan

expected_output

1 []

generated_code

null

code_valid_check

false

Next-up and Anticipated Challenges

- ChatAnthropic to Local Inference: Compute? Federated? GlobusCompute?
- Generalizability to multiple tables
 - Seed Text(s) vs Per-actor number of valid runs
 - Inter-actor agreement vs Critic actor
- Error traceability
 - Multi-actor state history with MLFlow?
 - Validator design and validation test cases
 - Critic design: Developer designed actor vs LLM intelligent actor

Agentic Stack

- Agentic middleware: LangGraph LangChain
- LLMs: Claude Models (CPU, API access)/ **Open-source Models (GPU)**
- Hardware: DGX Spark (NVIDIA GB10 GPU)
- Agentic middle-middleware: Academy
- Parallel Computing/Parallel Agent Management: Parsl
- Agent Message/Exchange Database: Redis (docker)
- Python Environment
- Experiment Tracking: MLFlow, Sqlite Database
- **SQLAlchemy Schema Import**