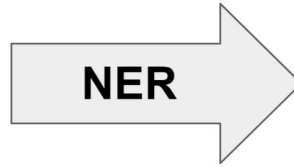


Book indices are a fast and easily accessible
source of evaluation data for NER

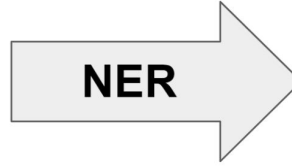
Named Entity Recognition



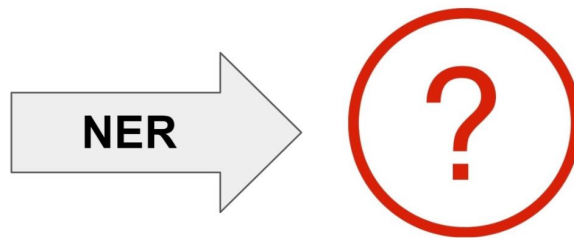
NER on Historical Documents



NER on Historical Documents

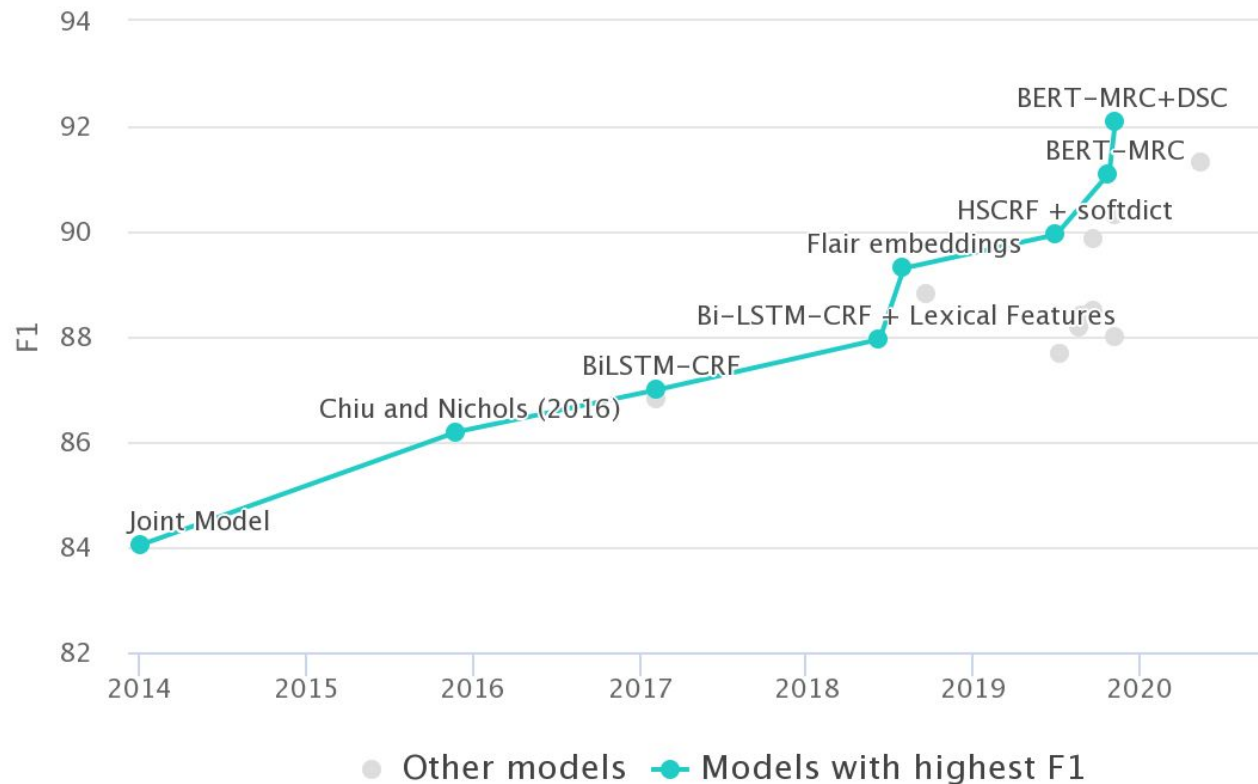


NER on Historical Documents



THE PASTORALISTS AND THE SHEARERS' UNIONS.	THE PASTORALISTS AND THE SHEARERS' UNIONS.
<p>An inquiry is (says the <i>Age</i> of Tuesday last) being instituted amongst members of the Pastoralists' Union in consequence of a number of communications having been received by individual station-owners from shearers in their employment expressing the reluctance with which they had struck work, and declaring that they had at the outset been forced to join the union against their will, and in dread of the ulterior consequences threatened in the event of their refusal. Furthermore, it has come to the knowledge of the Pastoralists' Union that many of these shearers obeyed the orders of the Shearers' Union to strike because they laboured under the impression that if they refused to do so they would be liable to have fines and penalties recovered from them by an action at law. So</p>	<p>An inquiry is (sajs tho Aqe of Tuesday last) being instituted amongst membois of thio Pastorahbts' Union in eoilibeception of a mimbi i of communications having been itceived bj individual station ownei s from slicieuis in their employment opycissing tho roluetancu with which tlity lind stiuck woik, and do clanng that they li id at the outset been foi cod to join the union against their will, and in dread of the ulterior consequences threatened in the event of their rofusa! Fuithermoro, it has como to tho knowldgo of tho Pastioahsts' Union that many of thioso shcaicrs othejed tho ofdeis of tho Shoaiers' Union to stuko because they laboured undoi the impression that if they refused to do so they would bo liable to have linos and penalties locovered from thom by an action at law.</p>

Novel NER Models



Evaluation

- It is essential to evaluate novel NER models on variety of settings.
- Creating ground-truth is expensive.



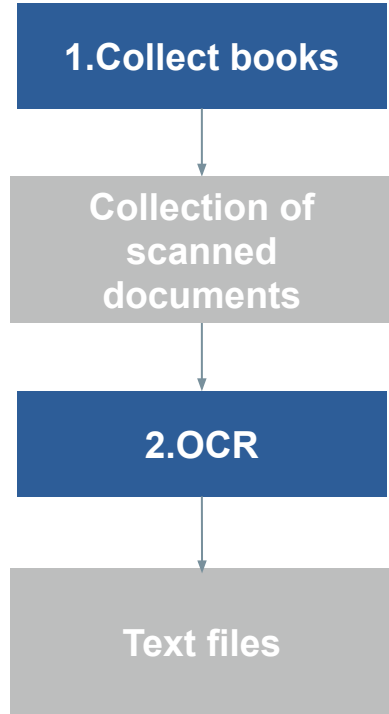
OCR-NER/L Pipeline

1. Collect books

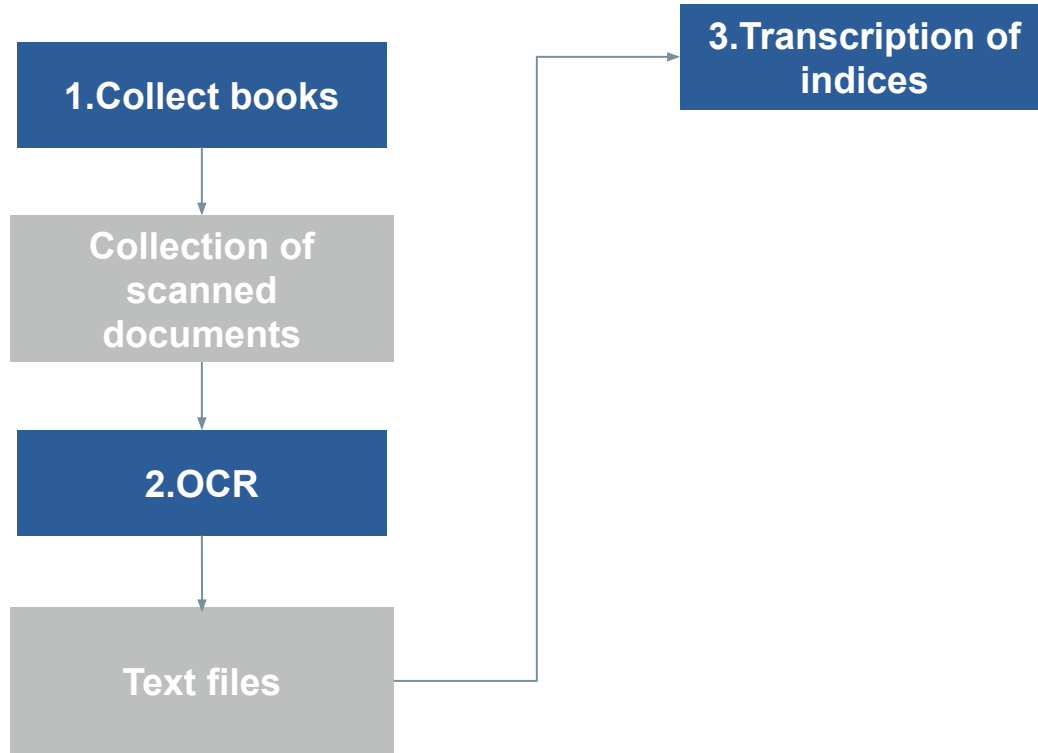


Collection of
scanned
documents

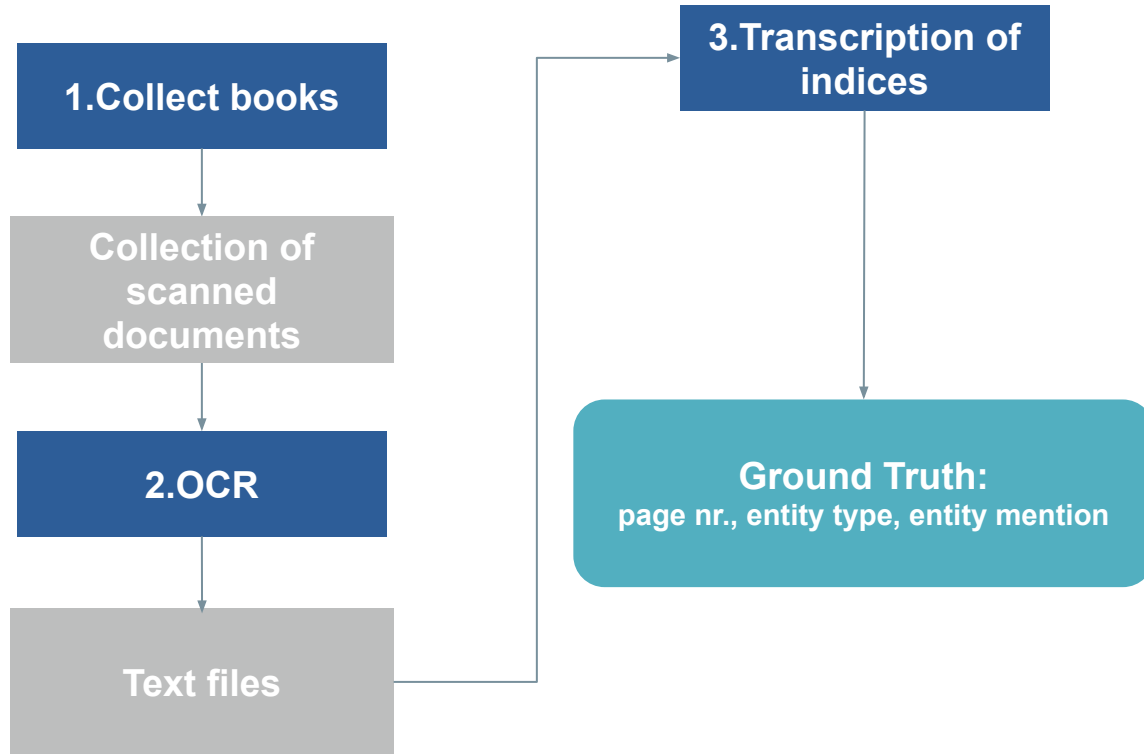
OCR-NER/L Pipeline



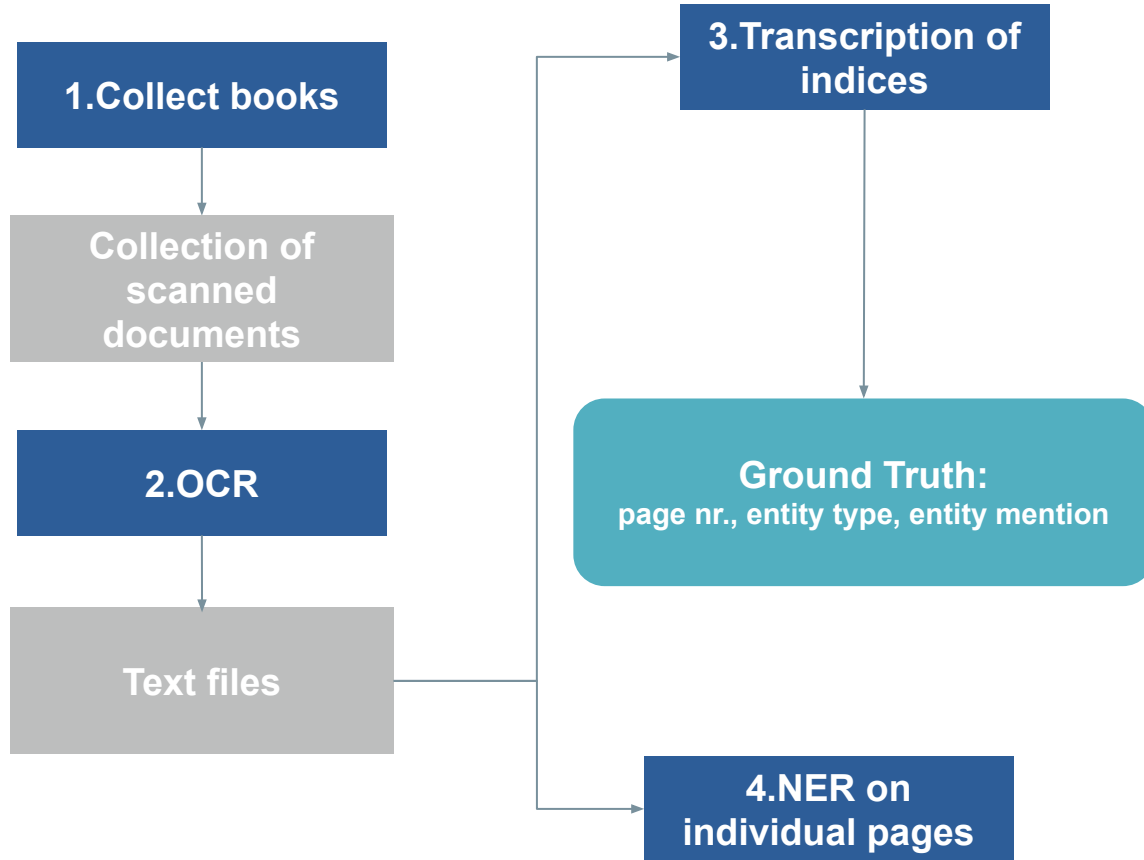
OCR-NER/L Pipeline



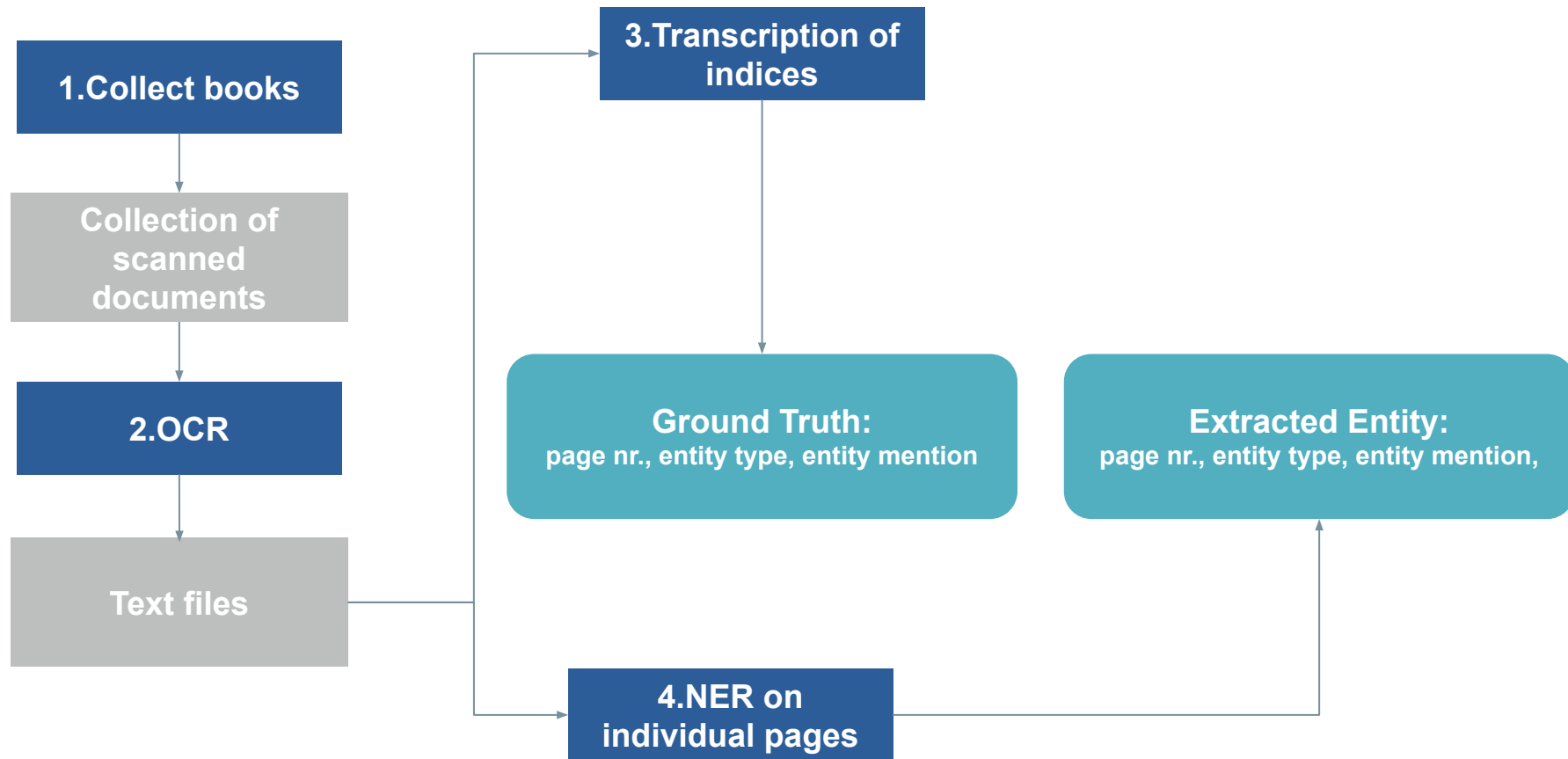
OCR-NER/L Pipeline



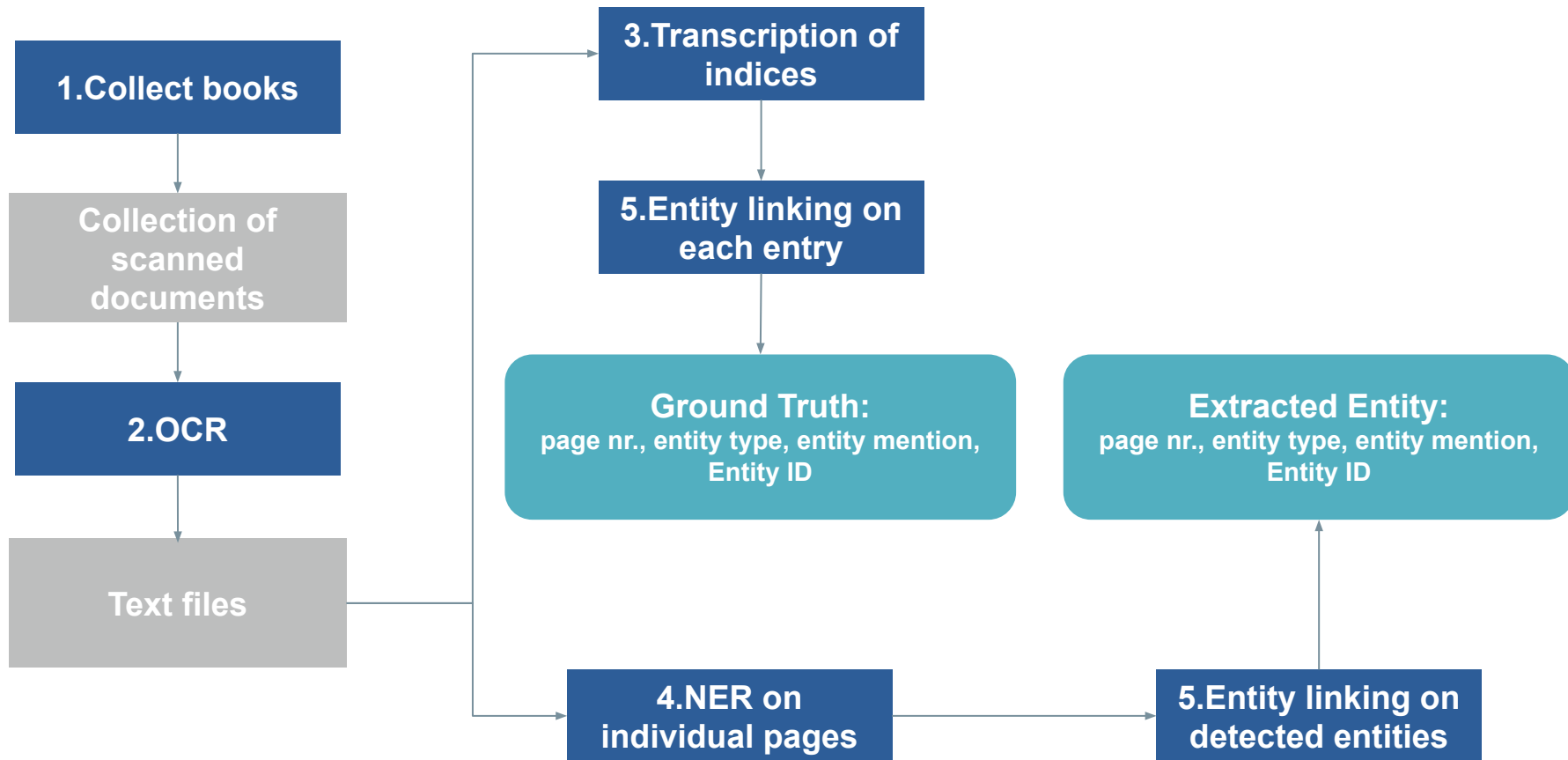
OCR-NER/L Pipeline



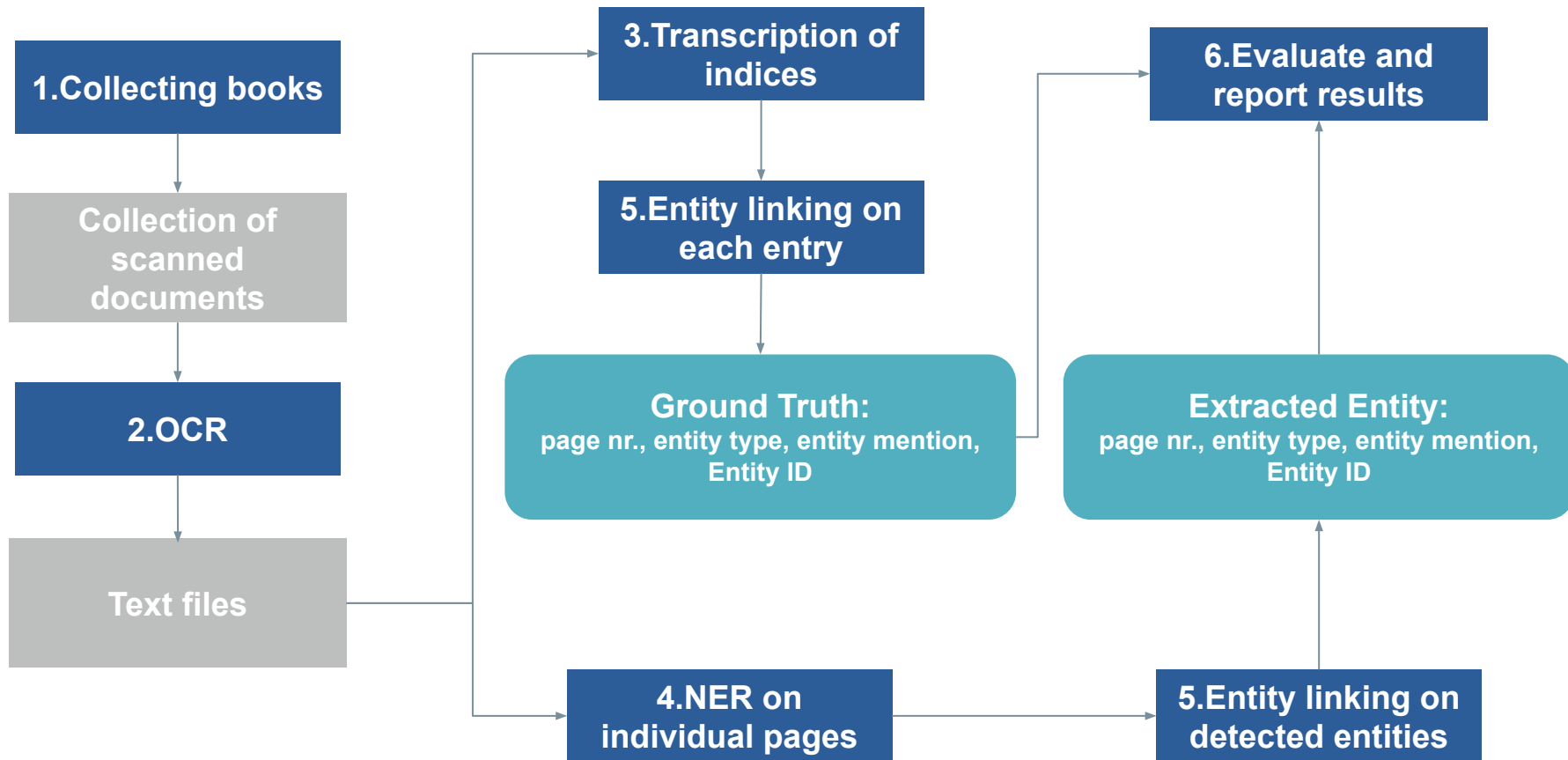
OCR-NER/L Pipeline



OCR-NER/L Pipeline



OCR-NER/L Pipeline



Historical Persian Astronomical Books

- Kraken OCR engine
- ParsNER

در حاله‌ها آسمان وزمین

نمایه مکانها

اول اقليم، ۱۲۹	آذربايجان، ۱۳۱
اهواز، ۱۳۰	ايسكون، ۱۳۵
ايله، ۱۳۳	ارمنيه، ۱۳۶
باورد، ۱۳۰	اسپيجاب، ۱۳۱
بحر محيط، ۱۳۵	اسفراين، ۱۳۱
بحيرات، ۱۳۶	اسكندريه، ۱۳۰
بحيره، ۱۳۶	اشروسنه، ۱۳۱
بحيره خوارزم، ۱۳۶	افرنجه، ۱۳۱
بحيره طبريه، ۱۳۶	افريقه، ۱۲۹، ۱۳۰
بخارا، ۱۳۱	اقليم اول، ۱۲۸، ۱۳۲
بدخشان، ۱۳۰	اقليم پنجم، ۱۳۱
بربر، ۱۲۹، ۱۳۰، ۱۳۳	اقليم دؤم، ۱۲۹
برده، ۱۳۱	اقليم سؤم، ۱۲۹
بست، ۱۳۰	اقليم ششم، ۱۳۱
بصره، ۱۳۰، ۱۳۴	اقليم هفتم، ۱۲۸، ۱۳۲
بطايع، ۱۳۶	انبار، ۱۳۰
بلاد روم، ۱۳۱	اندلس، ۱۳۱، ۱۳۲، ۱۳۴، ۱۳۵
بلاساغون، ۱۳۱	انطاكيه، ۱۳۱
بلخ، ۱۳۰	اوقيانوس، ۱۳۳، ۱۳۴

برجها و منزلهای شمالی برجهای شمالی شش‌اند، ج
و جنوبی کدامند. سنبه. زیرك منطقه البرو
از معدل النهار. و شش برج دیگر جنوبی‌اند.
آنچه اندر برجهای شمالی‌اند. و آن از اول شرطین
منزل از اول غفر تا بآخر بطن الحوت جنوبی‌اند
فلك مثل سطح منطقه البروج همه کو
کدامست. و بهر کره‌ای دایره‌ای کند
فلك مثل آن ستاره است که آن کره اوراست.

و قسم بالقیمت ه‌ای گفتیم این روز تالیث زهره و مشتری است. فی علامات
چنانک فی ج‌ای اختراق مریخ است و هم چنین فی علامت مشرق باشد و به
چنانک چون نبشته باشد دیخی آ فی گفتیم عطارد در مشرق پوشیده شود و اگر
یعنی پیدا آید.

باب هشتم: إدرا اجتماع و استقبال و خسوف

بر درازی کره ماهها سوی عرض اجتماع یا استقبال نوشته بود و اگر بروز
ساعات آن و طالع وقت و دهم برج او، و آن جزء کی شمس و قمر قدر و مسج
مقابل، و نبش برج از خانه آن نور بود کی نوبت او را باشد اگر بروز بود
شمس، و اگر شب بود از جایگاه قمر و این در جزء استقبال بود چی در اجتماع
یک برج باشد، و نمودار بطلمیوس را اعتماد برین جزء است. و الفاکسوف
آخر تقویم بر آن گونه کی اتفاق افتاد مقدار گرفتن، و ساعات و مدت و مکث و
آن برج کی اندرو بروز نبشته باشد و بدان کی قمر را خسوف اندر استقبال بود
کسوف اندر اجتماع، و رأس و ذنب یا بشان بروز مجتمع یا قریب بحدی معلوم.



Results

	True Positive	False Positive	False negative	Precision	Recall	F1 score
Person	3	136	91	.021	.031	.025
Location	51	101	223	.33	.18	.23

performance of the NER/-linking pipeline

Challenges

- Alleviating OCR error impact
- Language variation
- Persian is a low resource language