match\_redacted.txt - соотношение предобработанных источников и источников из указателя

match.txt / matched.json - соотношение НЕпредобработанных источников и источников из указателя

not\_matched.txt - источники из указателя, которые пока ни с чем не совпали

unique\_sources\_no match.txt - источники из текста, которые пока не были соотнесены с источниками из указателя

page\_out.txt - все теги <title> вне <bibl>

title\_out.txt - все теги <biblScope> вне <bibl>

sources\_freq.txt - в каких (и скольких) статьях можно найти каждый уникальный источник

unique\_sources.txt - список уникальных источников из текста