NTCIR-16 We Want Web with CENTRE (WWW-4)

www4org@list.waseda.jp

Zhumin Chu (Tsinghua University, P.R.C.)
Nicola Ferro (University of Padua, Italy)
Yiqun Liu (Tsinghua University, P.R.C.)
Maria Maistro (University of Copenhagen, Denmark)
Tetsuya Sakai (Waseda University, Japan)
Ian Soboroff (NIST, USA)
Sijie Tao (Waseda University)

http://sakailab.com/www4/

Version 20210325

Introduction

- WWW-4 as an adhoc English web search task!
- Web search is not a solved problem! The new deep learning approaches have given us new open questions! IR researchers should find out what works and what doesn't, and why!
- We will introduce a new target web corpus based on Common Crawl data!
- Missions:

- Quantify the progress in adhoc web search algorithms across the NTCIR rounds
- Ensure replicability and reproducibility

What is CENTRE?

Consistent with ACM's original definitions

- Replicability: Team A reports results on Data D; Team B obtains the same results on D.
- Reproducibility: Team A reports results on Data D; Team B obtains similar results on another data D'.
- In reality, the above are hard to achieve. CENTRE = CLEF/NTCIR/TREC Reproduciblity was a meta-conference effort to address this.
- We had a CENTRE at TREC 2018, CLEF 2018, 2019, and NTCIR-14. It lives on as part of the WWW task.

Run types

- **REV (revived) run – for University of Tsukuba only** Tsukuba's KASYS-E-CO-NEW-1 was a top performing run at the NTCIR-15 WWW-3 English subtask. As described in their paper (KASYS at the NTCIR-15 WWW-3 Task), this is a BERT-based run. For the WWW-4 task, Tsukuba will use exactly the same system to process the new WWW-4 topics on the new target corpus. Hence this "revived" run represents the current state of the art.
- **NEW (standard adhoc) runs** This is a good old adhoc web search task. Please process the NTCIR-16 WWW-4 test topics with your proposed system. Your run will be compared to not only other runs but also the state of the art from NTCIR-15 WWW-3, namely the REV run (see above). Can you outperform it and become the new state of the art?
- **REP (reproduced) runs** Reproducibility is crucial for advancing the state of the art as a research community. Please read the KASYS paper and try to reproduce their system! We will evaluate REP runs in terms of how similar they are to the REV run, and whether the effects over the BM25 baseline are preserved.

Topic set size design

http://link.springer.com/content/pdf/10.1007%2Fs10791-015-9273-z.pdf (open access)

- We obtained variance estimates that are probably more reliable than those from the previous WWW rounds, with as many as 160 topics and 36 runs. From the 160x36 score matrices, we obtained 0.0284 for nERR (largest among the four measures), and 0.00716 for iRBU (smallest among the four measures). According to topic set size design for the t-test under Cohen's five-eighty convention,
  - We need 44 topics for a minimum detectable difference of 0.1 in terms of nERR;
  - We need 45 topics for a minimum detectable difference of 0.05 in terms of iRBU.
- Hence we will construct 50 topics to achieve at least 80% power for the above minimum detectable differences.

Evaluation measures (same as WWW-3)

- Retrieval effectiveness: nDCG@10, Q@10, ERR@10, iRBU@10
- Replicability and reproducibility:
- Effect Ratio (for replicability and reproducibility)
- Correlation of per-topic Δ's (replicability)
- RMSE of per-topic Δ's (replicability)

For details, see the NTCIR-14 CENTRE overview:

http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings14/pdf/ntcir/01-NTCIR14-OV-CENTRE-SakaiT.pdf

https://www.slideshare.net/TetsuyaSakai/ntcir14centreoverview

Important dates (UTC+9)

April 2021	Announcement of the new corpus/Task registrations open
Oct 1 2021	Topics released; task registrations due
Nov 15 2021	Run submissions due
Dec 2021-Jan 2022	Relevance assessments
Feb 1, 2022	Evaluation results released
Feb 1, 2022	Draft task overview paper released
Mar 1, 2022	Draft participant paper submissions due
May 1, 2022	All camera-ready paper submissions due
Jun 2022	NTCIR-16 Conference in NII, Tokyo, Japan