NTCIR-15 We Want Web with CENTRE (WWW-3) Task Definition

www3org@list.waseda.jp
Zhicheng Dou (Renmin University of China, P.R.C.)
Nicola Ferro (University of Padua, Italy)
Yiqun Liu (Tsinghua University, P.R.C.)
Maria Maistro (University of Padua, Italy)
Jiaxin Mao (Tsinghua University, P.R.C.)
Tetsuya Sakai (Waseda University, Japan)
Ian Soboroff (NIST, USA)
Sijie Tao (Waseda University)
Zhaohao Zeng (Waseda University)
Yukun Zheng (Tsinghua University)

Version 20190818

Introduction

- WWW-3 is an adhoc web search task for Chinese and English.
- All WWW-3 participating systems will process both WWW-2 (80) and WWW-3 (80) topic sets.
- Three run types for the English subtask:

REV (revived) runs: a top team from the WWW-2 task rerun two of their systems.

REP (replicated/reproduced) runs: other teams try to copy the above two WWW-2 systems.

NEW runs: regular adhoc and any other runs.

- The Chinese subtask will probably have NEW runs only (not finalised).

Why are we still doing web search?

- Web search is not a solved problem! Queries are poor representations of information needs; we're still doing keyword matching! We don't know what's NOT retrieved! We need to know what works and what doesn't, and keep improving!
- With the advent of deep learning approaches, we have even more open questions in web search than before!
- Progress should be monitored across years! WWW-3 will be directly compared to WWW-2!

What is CENTRE?

- 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 NTCIR-15 WWW-3.

Chinese subtask

- Input: 80 WWW-2 topics + 80 new WWW-3 topics
- Target corpus: SogouT-16 (same as WWW-2)
- Output: A TREC-style run file
- Researchers can also participate without indexing the corpus, by reranking the baseline runs provided by the organisers.
- As an external resource, a large query log Sogou-QCL can be utilised so that researchers can try intensive machine learning approaches
- For more info on the corpora and resources, see the WWW-2 overview paper:

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

English subtask

- Input: 80 WWW-2 topics + 80 new WWW-3 topics
- Target corpus: clueweb12-B13 (same as WWW-2)
- Output: A TREC-style run file
- Researchers can also participate without indexing the corpus, by reranking the baseline runs provided by the organisers.
- Three run types: REV runs (Tsinghua University only), REP runs (any teams that want to try replicating/reproducing the Tsinghua runs), NEW runs (for those who want to try their own algorithms).

Target runs

- At NTCIR, CENTRE focuses on replicating/reproducing the effect of a method over another. At NTCIR-15, we chose the following pair to address a simple question: does a SOTA learning-to-rank run outperform BM25 regardless of experimental conditions (i.e., people and data)?
- A-run (advanced run)

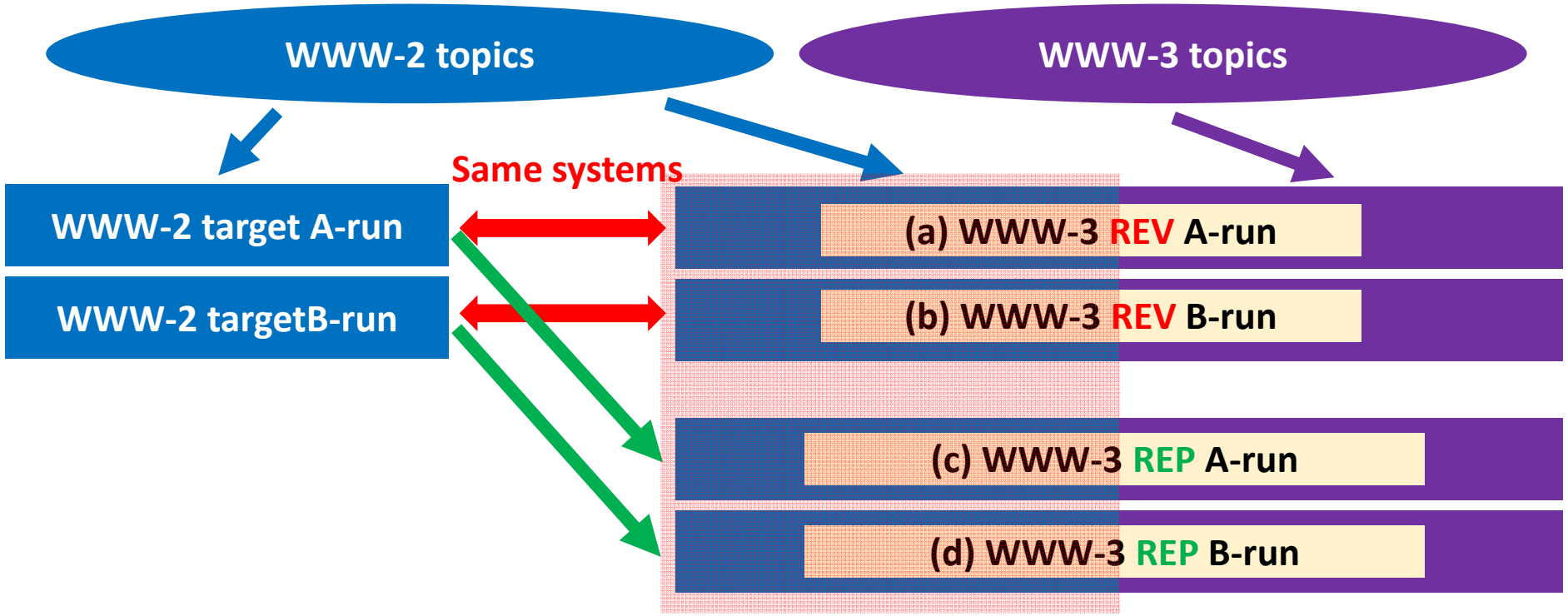
THUIR-E-CO-MAN-Base-2 (LambdaMART)

- B-run (baseline run)

THUIR-E-CO-PU-Base-4 (BM25)

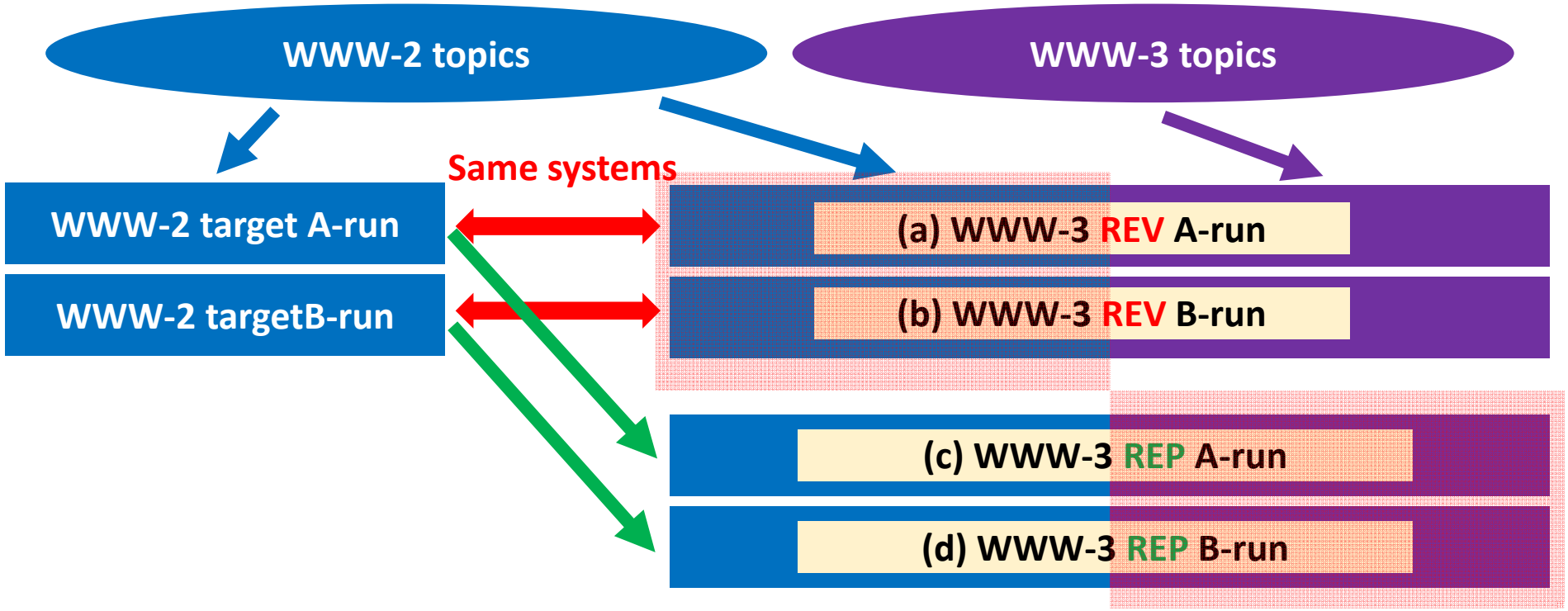
Details of these Tsinghua runs can be found in their NTCIR-14 paper:
http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings14/pdf/ntcir/03-NTCIR14-WWW-ZhengY.pdf

Testing replicability



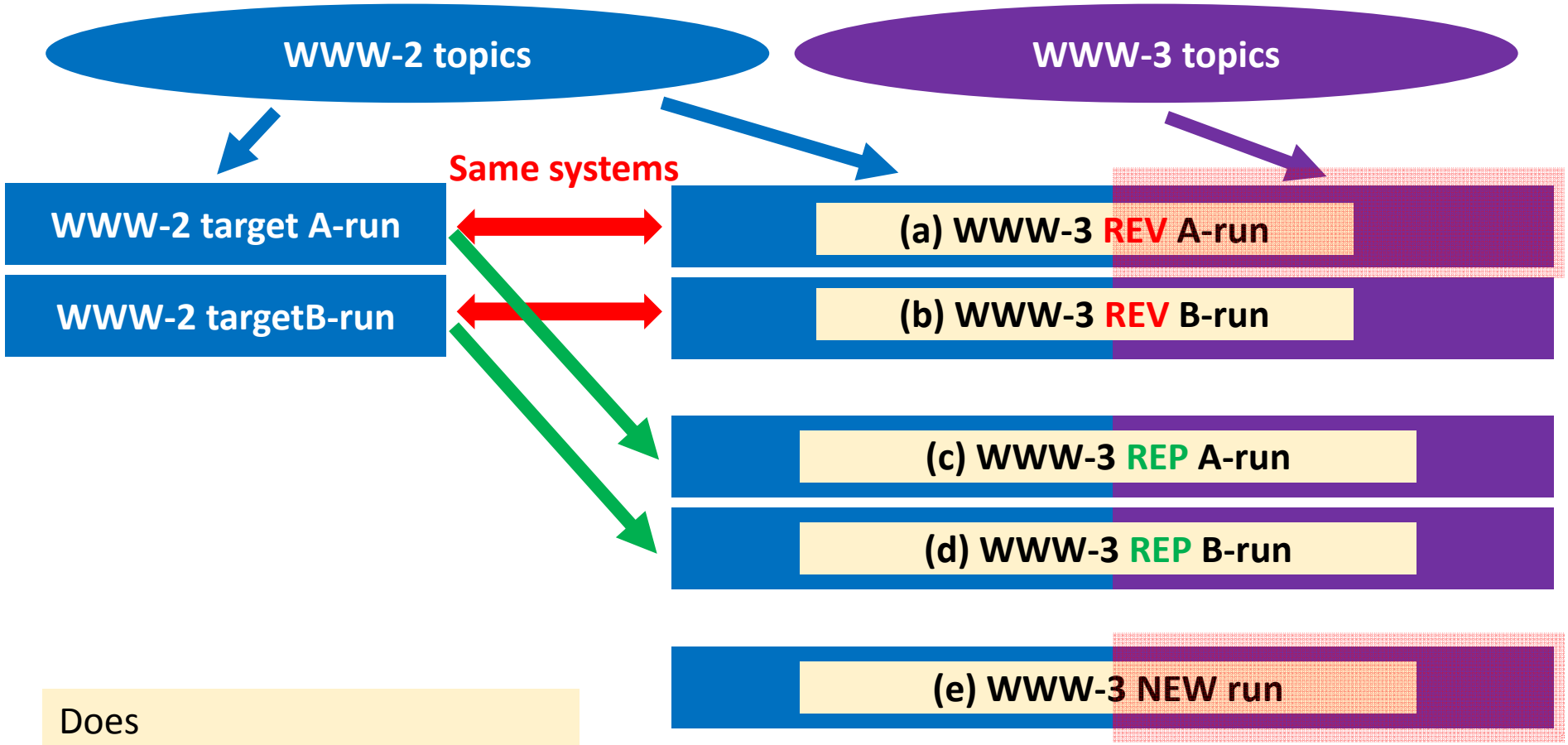
Does
(a)>(b) ⇒ (c)>(d)
hold on the WWW-2 topic set?

Testing reproducibility



Does
(a)>(b) on the WWW-2 topic set
⇒ (c)>(d) on the WWW-3 topic set
hold?

Measuring progress



Does
(e) > (a)
hold on the WWW-3 topic set?

Note that (e) had access to WWW-2 qrels, unlike (a)

Evaluation measures

- Retrieval effectiveness: nDCG@10, Q@10, ERR@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