**Procedure of online evaluation**

* Due to operational constraints (difficult to incorporate user study for all testing papers), only 20 sub-domains of computer science were selected as a testing dataset in our experiment.
* A total of 160 seed papers are chosen manually from 20 sub-domains: information retrieval (IR), image processing (IP), security (SC), wireless sensor network (WSN), machine learning (ML), software engineering (SE), computer vision (CV), artificial intelligence (AI), data mining (DM), theory of computation (TC), databases (DB), human-computer interaction (HCI), algorithms and theory (AT), natural language processing (NLP), parallel and distributed systems (PDS), worldwide web (WWW), web semantics (WS), computer architecture (CO), compiler design (CD) and multimedia (MM).
* There are 78 researchers with expertise in the mentioned sub-domains are provided with input and output of our recommender system where for each paper, 15 venues are recommended. Out of 78 researchers, 29 evaluated 3 papers each, 24 researchers evaluated 2 each, and the rest 25 were evaluated by 25 researchers. All the experts were identified from academia with a minimum of 3 years of research experience.
* Most were having a Ph.D. except few research students and research assistants who were pursuing a Ph.D. with bachelors' or masters' degree in science or technology. The experts or researchers were so chosen that their active areas of research perfectly match with the topics of seed papers. Among 78 researchers, there were 16 professors, 15 associate professors, 23 assistant professors, 18 senior research students, and the remaining 6 were research assistants.
* The experts check the titles, abstracts, authors, year of publication, and recommended venues of the papers. An expert assigns an appropriate relevance value ($r$) to each recommended venue as she deems the quality of the match between the scope of the recommended venue and the topic of the seed paper as below.
* 2-perfectly matching , 1-partial matching , 0-otherwise
* However, as precision is defined for binary relevance only, during precision score computation, relevance grade $2$ is only considered relevant, and both relevance grade $1$ and $0$ non-relevant.