Table 4.2. Subareas of Computer Science selected from Microsoft classification.

Subarea	Abbreviation	Seed venues
Algorithm	Alg	SODA, Algorithmica
Artificial intelligence	AI	IJCAI, AI
Bioinformatics	Bio	BIBM, Bioinformatics
Computer graphics	CG	SIGGRAPH, TCVG
Computer network	CN	INFOCOM, TON
Computer security	CS	CCS, TISSEC
Computer vision	CV	CVPR, IJCV
Data mining	DM	KDD, SIGKDD
Database	DB	SIGMOD, TODS
Distributed computing	DC	ICDCS, TPDS
Human-computer interaction	HCI	CHI, TOCHI
Information Retrieval	IR	SIGIR, TOIS
Machine learning	ML	ICML, JMLR
Natural language processing	NLP	EMNLP, COLING
Operating system	OS	SOSP, SIGOPS
Parallel computing	PC	IPPS, TPDS
Programming language	PL	PLDI, TOPLAS
Speech Recognition	SR	INTERSPEECH, TCOM
Theoretical computer science	TCS	STOC, SIAMCOMP
World Wide Web	WWW	WWW, WS

for the connect application of P-score to subareas. We avoided using all 37 subareas because we found some of them difficult to characterize. While our subset of 20 CS subareas is not perfect or exhaustive, it is detailed enough to allow gaining insights into the scene of research in CS in Brazil, which would not be possible to obtain otherwise. In Appendix B, we present other classifications of CS subareas, according to reliable sources.

4.3 Venues Ground-Truth

To evaluate the effectiveness of normalized P-scores from Equation (3.5) on the task of finding venues in a subarea, we considered as ground-truth the opinion of experts. Specifically, we asked reputable CS researchers and their graduate students, working on subareas of IR, DB and Data Mining (DM) to assess the relevance to their subarea of venues included in a pre-selected list. This list consists of the venues at the top 50 positions in the P-score ranking when we use as seeds two publication venues only: a journal and a conference closely associated with that subarea. For examples of seeds, see Table 4.3.