CS 5500 Course Project
Phase I (Assigned: January 24, Due: February 3)

Frank Tip and Mike Weintraub

High Level Objective

Finding a diverse set of suitable and available reviewers for judging conference or journal paper submissions is a challenging and time-consuming problem for scientific publishers. The objective of the course project is to create a system that leverages existing conference and journal repositories to propose appropriate candidates who can adequately review submissions to a Program Committee Chair or Associate Editor.

Background

Publications

Computer scientists publish papers about their research in conference proceedings and journals. All papers have bibliographic information: a title, one or more authors, and other identifying characteristics depending on whether the publication appeared in a journal or conference. For example, conferences often have nicknames - usually an acronym over its full name (e.g., "OOPSLA" stands for the ACM SIGPLAN Conference on Object-Oriented Programming, Systems, Languages, and Applications). Journal names are abbreviated similarly (e.g., "ACM TOPLAS" stands for the journal "Transactions On Programming Languages And Systems", which is published by the Association for Computing Machinery. Journal articles typically appear in a specific volume and issue. Both conference and journal publications cover a range of page numbers, and have a specific year of publication.

Accessing Publication Data

Bibliography information is available on-line; two of the larger libraries are the DBLP Computer Science Biography (http://dblp.uni-trier.de/) and the ACM Digital Library (http://dl.acm.org/). Within these sites, it is possible to search for publications in many ways. For example, one can search a specific conference for a publication by a specific author. For example, the papers that were published in the proceedings of OOPSLA 2016 (the set of papers published at this conference) can be found here:

>   http://dblp.uni-trier.de/db/conf/oopsla/oopsla2016.html#0001GJSSTC16 [DBLP]

>   http://dl.acm.org/citation.cfm?id=2983990&picked=prox [ACM DL]

Peer Review process

An important part of the publication process is *peer review*. Authors submit papers they want published in a particular conference or journal. An organization – the conference or journal – has these submissions reviewed by one or more peers, i.e., experts in a field of study. This involves a great deal of work. For example, OOPSLA receives a few hundred paper submissions each year,

from which only 20-30% are selected for publication. For a conference such as OOPSLA, reviewing is done by a program committee (PC) of experts, which is managed by a program committee chair (PC chair). A new PC is set up each year, for each instance of a conference. Each program committee member is assigned a subset of that year's submitted papers, and then writes a review for each one. When all reviews of all submissions are written, the program committee chair convenes the program committee for a meeting to discuss each submission and its reviews, and then makes a decision to accept or reject it.

With journals, the peer review process is more straightforward. Here, instead of forming a PC for each conference, a journal has an editorial board. This editorial board consists of associate editors, subject experts who oversee the reviewing of papers submitted to the journal, and is led by an editor in-chief who is ultimately responsible for making accept/reject decisions. Associate editors serve a fixed term on the editorial board (e.g., for three years), and their term can be renewed. The review process is similar to the conference submission review process. For each submitted paper, an associate editor solicits reviews from appropriate subject matter experts. The associate editor collects these reviews and then makes an accept or reject recommendation to the editor-in-chief, who in turn, makes the final decision.

Logistics

For both conference and journal reviews, finding suitable reviewers is hard. Program chairs and associate editors often rely on their personal expertise and professional networks to identify suitable reviewers. However, relying on this approach too much can lead to real or perceived bias in the decision-making process. Another approach is to use leading subject matter experts who are famous or well-known in their field of study. While identifying such people can be easy, they often decline reviewing requests because they are too busy. Program committee chairs also need to consider other constraints. For example, it is considered desirable to have a reasonable balance of junior and senior people on a committee, to have a mix of committee members from different geographies (North America/Europe/Asia), and to have a reasonably diverse set of committee members. Furthermore, organizational guidelines may prohibit a person from serving on the program committee for the same conference for more than two consecutive years.

For these reasons, program chairs and associate editors need to take steps to identify suitable committee members and reviewers. This may include identifying people who published papers on similar topics in conferences with similar scope. For example, to find a suitable reviewer for an OOPSLA conference submission on "pointer analysis," a search on DBLP might reveal candidates who have published on this topic in previous instances of OOPSLA or ECOOP (European Conference on Object-Oriented Programming). Both of these conferences focus on object-oriented programming, so people who publish in one of them may be suitable reviewers for the other. It is also useful to know if potential candidates have previously served on PCs for conferences with similar scope, to satisfy any constraints on committee composition. Information about program committees is generally available from conference web sites. Further suggestions for possible reviewers could be mined from the list of citations in a paper, and such information can be accessed from the ACM Digital Library.

Project Goals

The goal of this project is to design, implement, test, and evaluate an application that assists program committee chairs and associate editors with identifying suitable candidates for their committees. The system you are building should provide an interactive user-interface to execute queries against publication data from sources such as DBLP and ACM DL. For example, the application should be able to answer queries such as:

- Identify the names of authors who published more than one paper in OOPSLA since the year 2010.
- Identify the names of authors who published at least one paper with the words "pointer" and "analysis" in the title.
- Identify the names of authors who published at least two papers in OOPSLA or ECOOP and did not serve on the committee during the last two years.
- Identify authors who have "similar" profiles to a given author.

The implementation of the system will take place in four phases, which will involve requirements gathering, design, implementation, refactoring, and testing.

Deliverables for Phase 1

The first phase of the project is to be done individually. There is no collaboration allowed! In this phase, you are required to write 5-10 use cases that describe the system's functionality. The clients (aka your instructors) will be present in class on January 27 for a Q&A session. Your deliverable must be submitted as a PDF file in BlackBoard by February 3 at 23:59pm. Your submission must be no longer than five (5) pages.