COMP 3311
Database Management Systems

Lab 9

Task 3 Logins and Screenshots

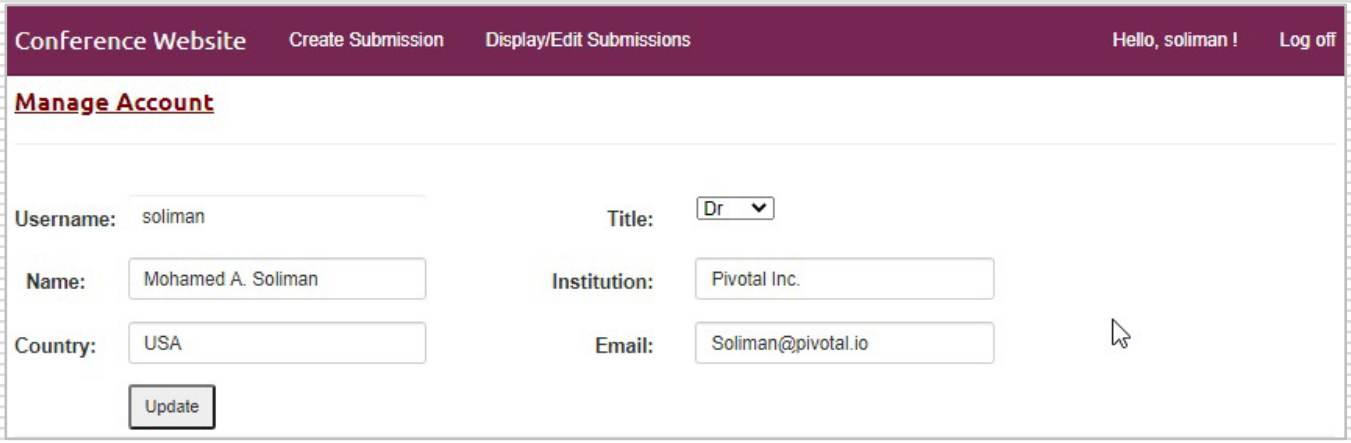
Logins Available for Task 3

PC Chair
lock

PC Members
chen
crestan
han
jag
larson
lock
ooi
pantel
rudens
yuj

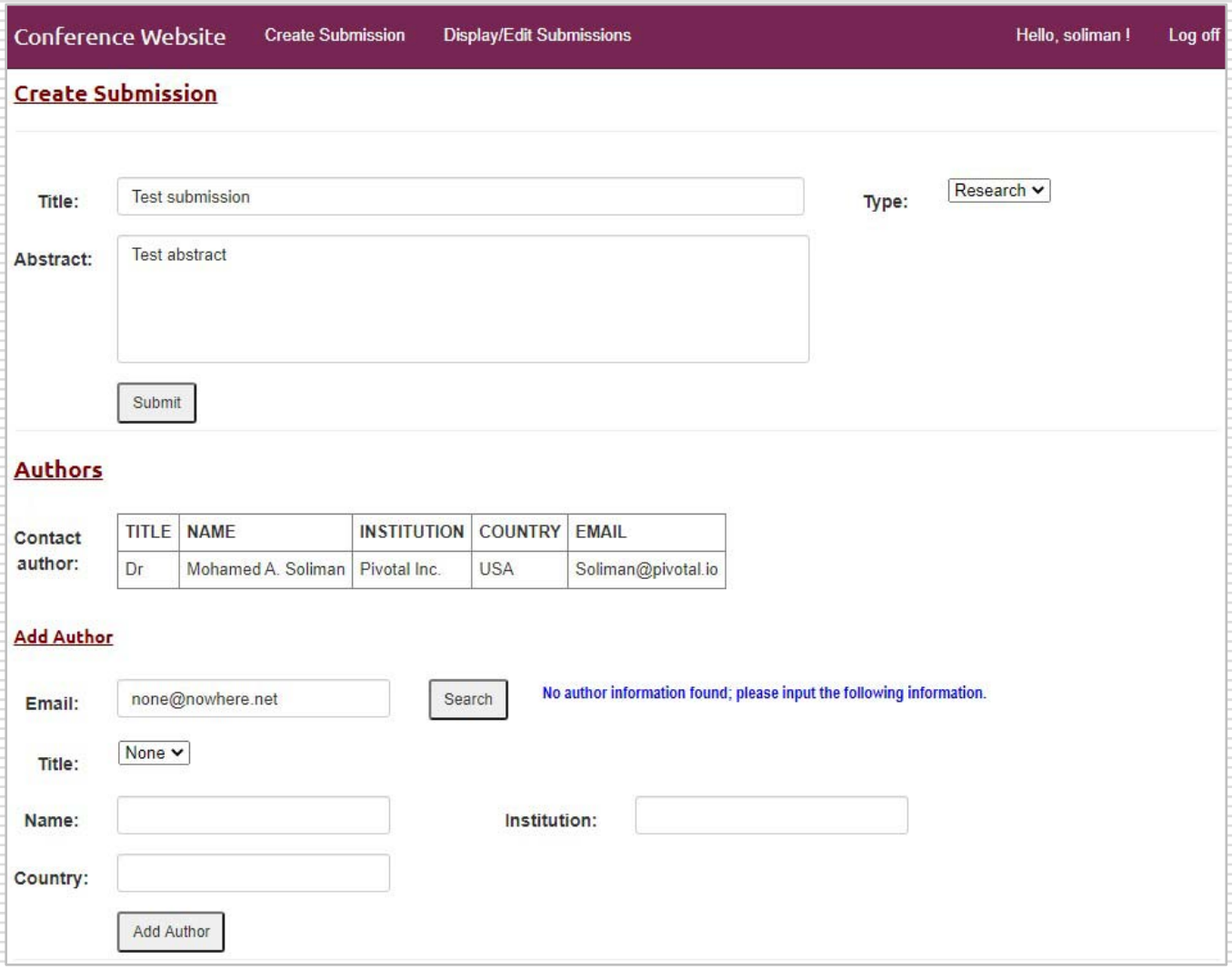
Contact Author
cafarella
crestan
fan
han
krishna
madden
molina
rec
soliman
swart
varma
zadoro

All Users: Manage Account

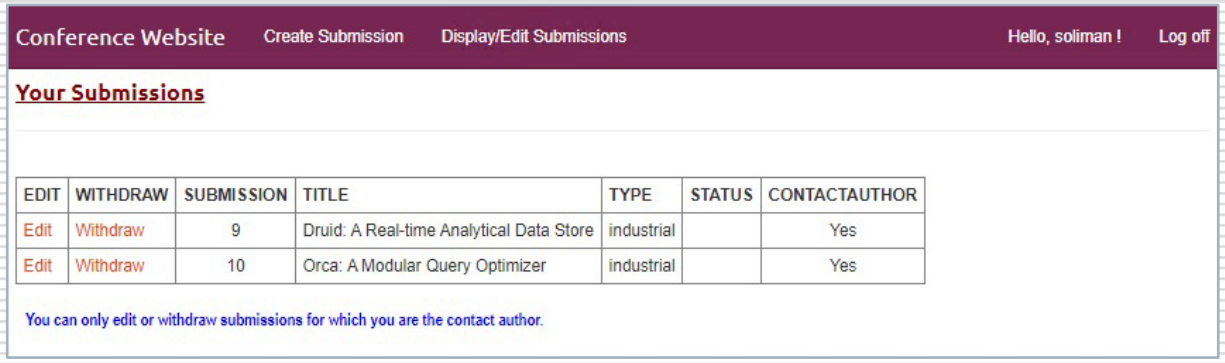


This page is accessed by selecting the "Hello, <usename>!" link in the menu bar.

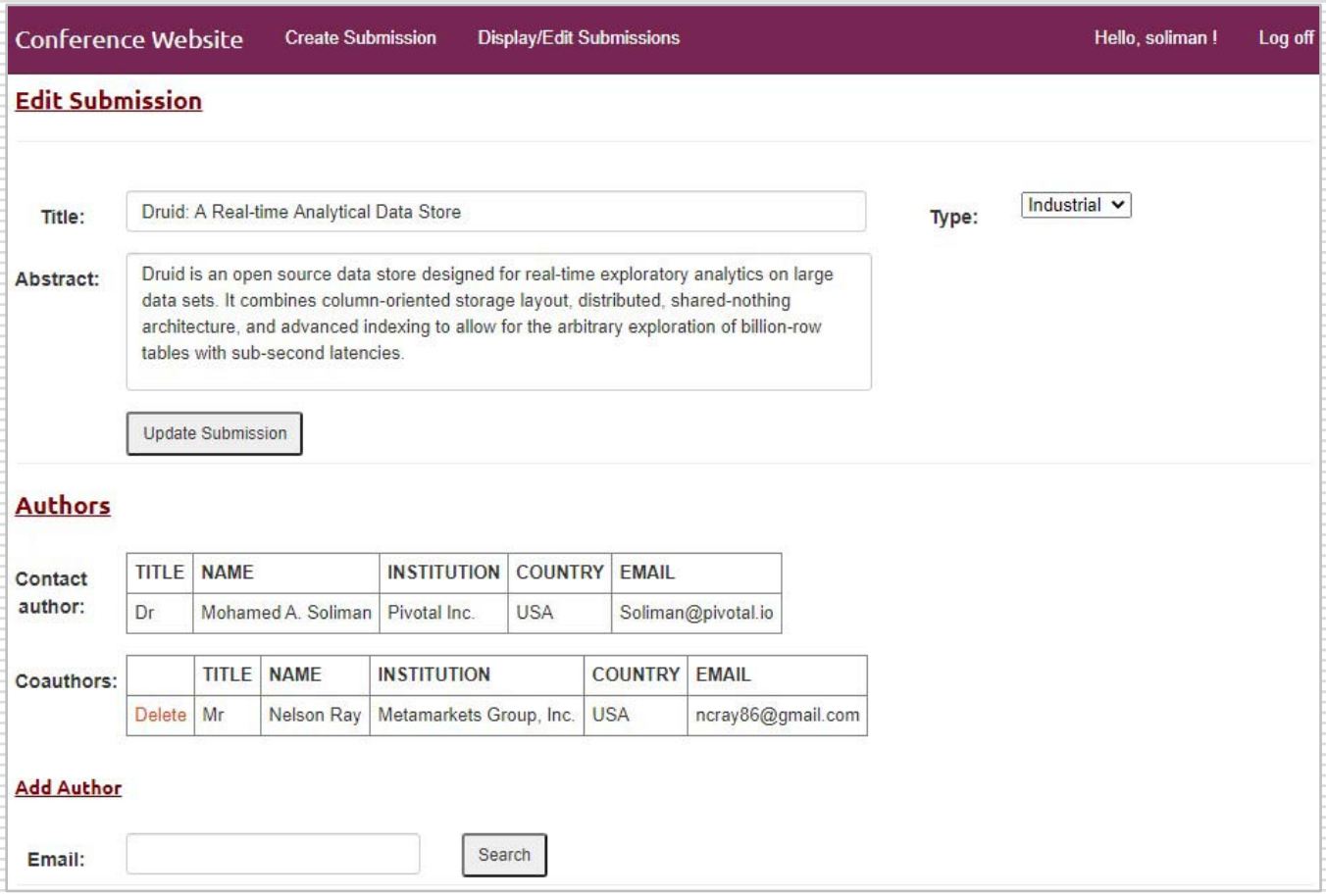
Author: Create Submission



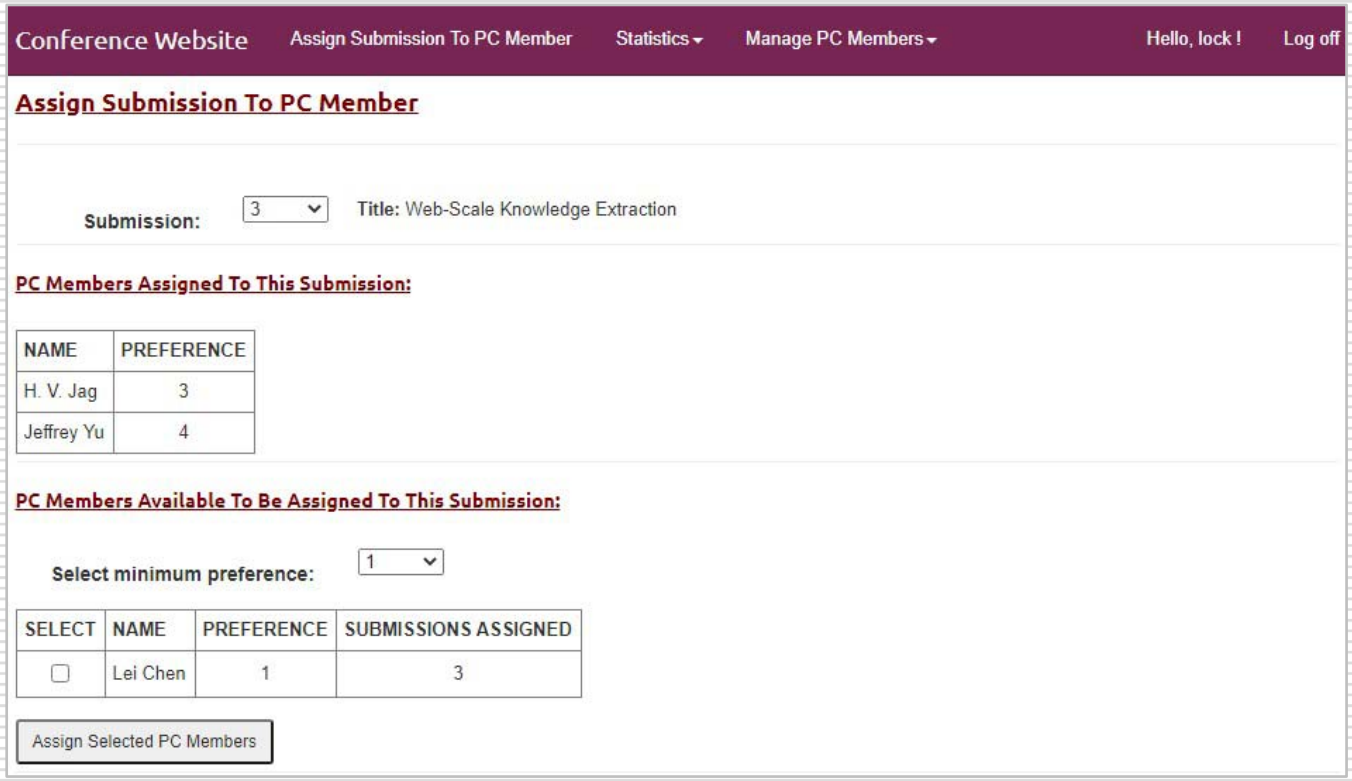
Author: Display/Edit Submission



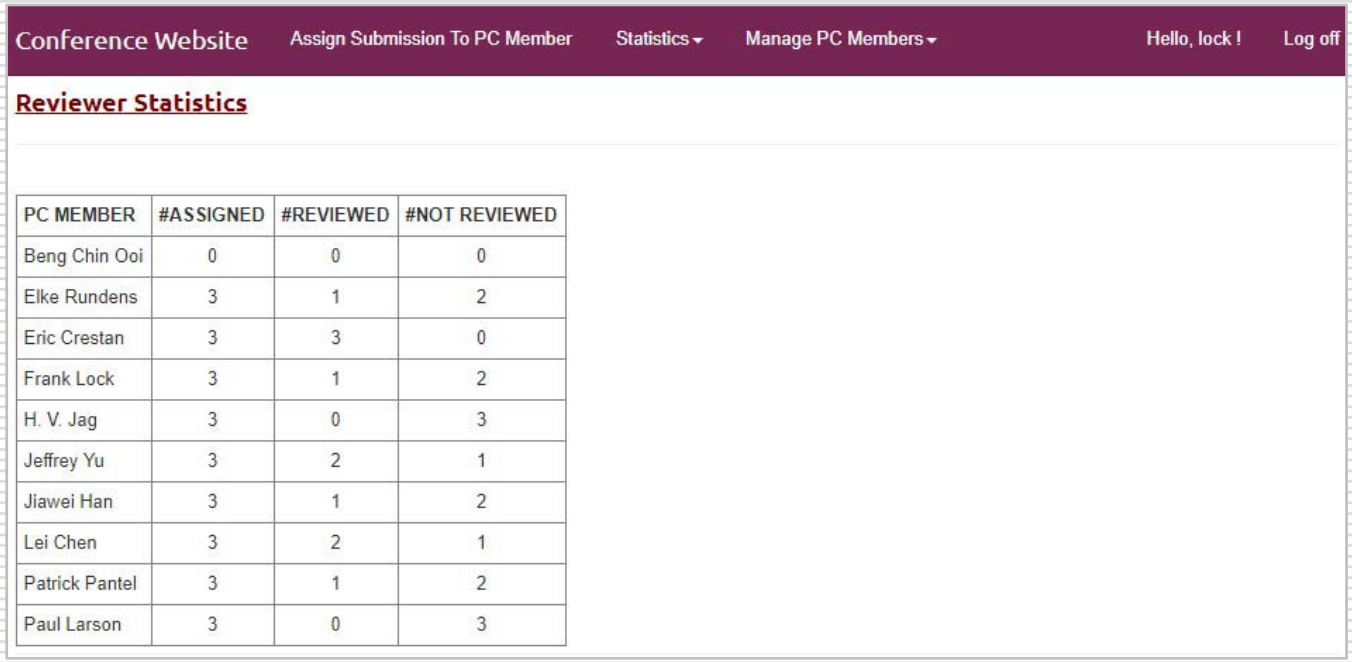
Author: Edit Submission



PC Chair: Assign Submission To PC Member



PC Chair: Reviewer Statistics



PC MEMBER	#ASSIGNED	#REVIEWED	#NOT REVIEWED
Beng Chin Ooi	0	0	0
Elke Rundens	3	1	2
Eric Crestan	3	3	0
Frank Lock	3	1	2
H. V. Jag	3	0	3
Jeffrey Yu	3	2	1
Jiawei Han	3	1	2
Lei Chen	3	2	1
Patrick Pantel	3	1	2
Paul Larson	3	0	3

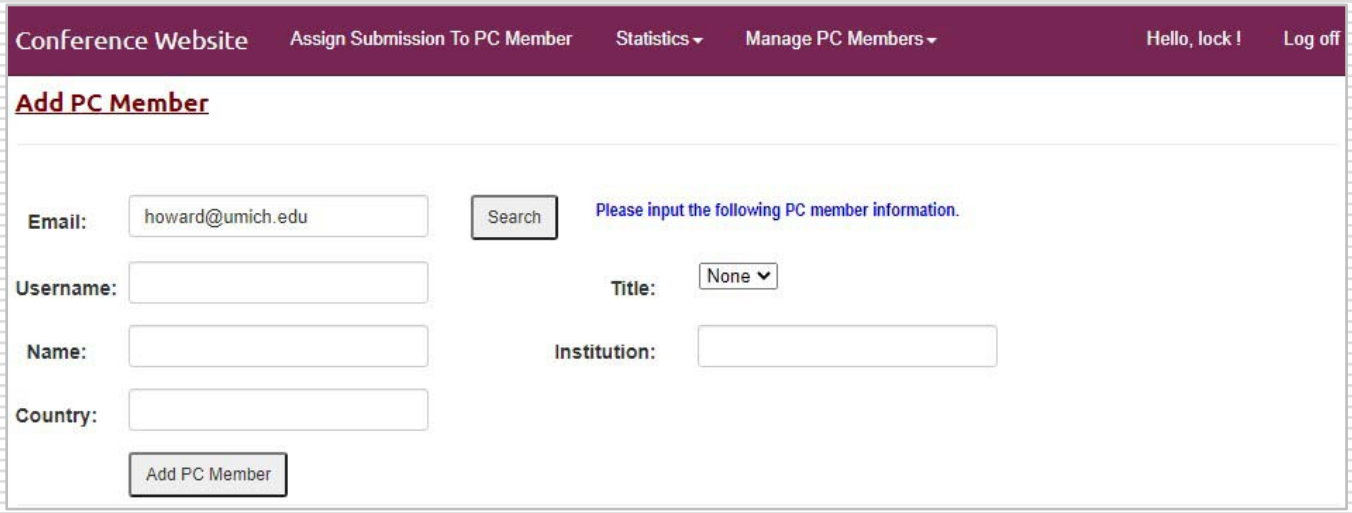
PC Chair: Submission Statistics

Conference Website | Assign Submission To PC Member | Statistics | Manage PC Members | Hello, lock ! | Log off

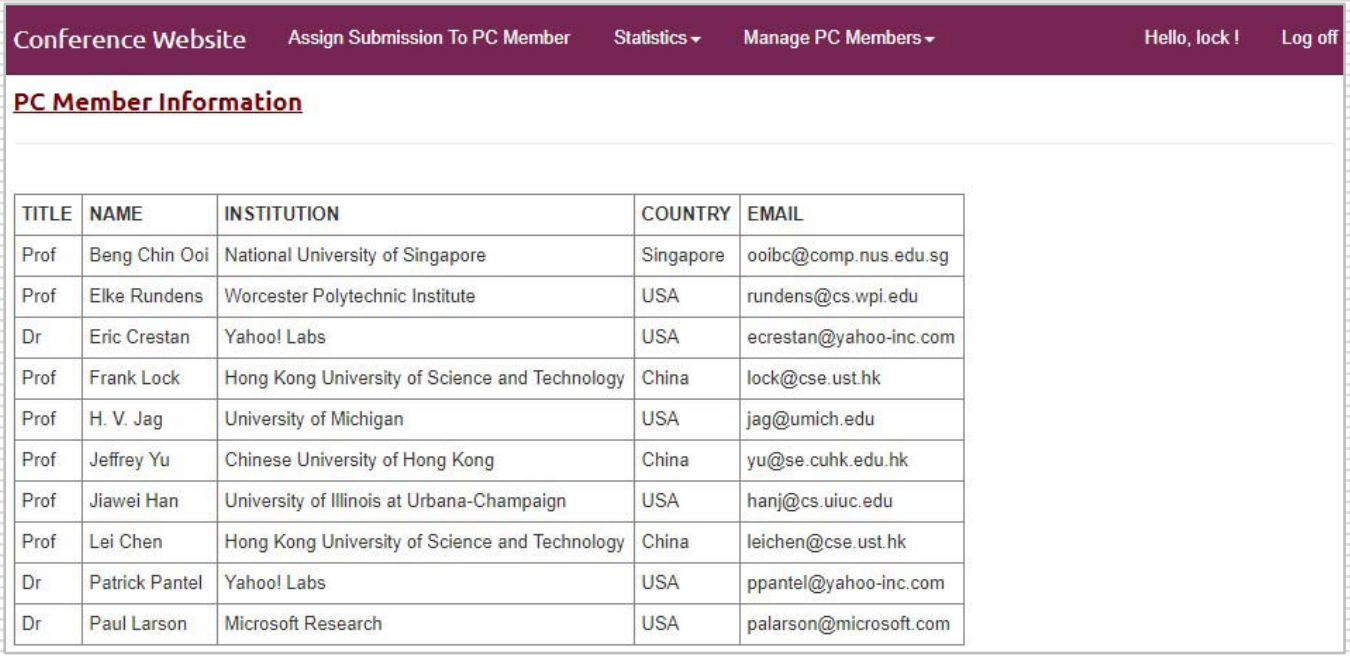
Submission Statistics

SUBMISSION	TITLE	TYPE	CONTACT AUTHOR	#AUTHORS	#REVIEWERS	#COMPLETED	STATUS
1	Example-Driven Schema Mapping	research	Michael J. Cafarella	2	3	3	reject
2	Optimal Schemes for Robust Web Extraction	research	Hector Molina	4	3	3	accept
3	Web-Scale Knowledge Extraction	research	Eric Crestan	2	2	0	
4	Efficient Fusion of Historical Data	research	Vladimir Zadoro	2	3	3	
5	CETR - Content Extraction via Tag Ratios	research	Jiawei Han	3	2	0	
6	Towards User-Friendly Entity Resolution	research	Rajasekar Krishna	2	3	2	
7	TsingNUS: A Location-based Service System	demo	Ju Fan	1	3	0	
8	A Java Stream Computational Model for Big Data	industrial	Garret Swart	1	3	0	
9	Druid: A Real-time Analytical Data Store	industrial	Mohamed A. Soliman	2	2	0	
10	Orca: A Modular Query Optimizer	industrial	Mohamed A. Soliman	1	3	0	withdrawn
11	Amoeba: A Shape Changing, Big Data Storage System	demo	Samuel Madden	4	0	0	
12	Snuba: Automating Weak Supervision to Label Training Data	research	Paroma Varma	2	0	0	
13	Training with Natural Language	industrial	Christopher Re	1	0	0	
14	SystemML: Declarative Machine Learning on MapReduce	research	Rajasekar Krishna	1	0	0	withdrawn

PC Chair: Add PC Member

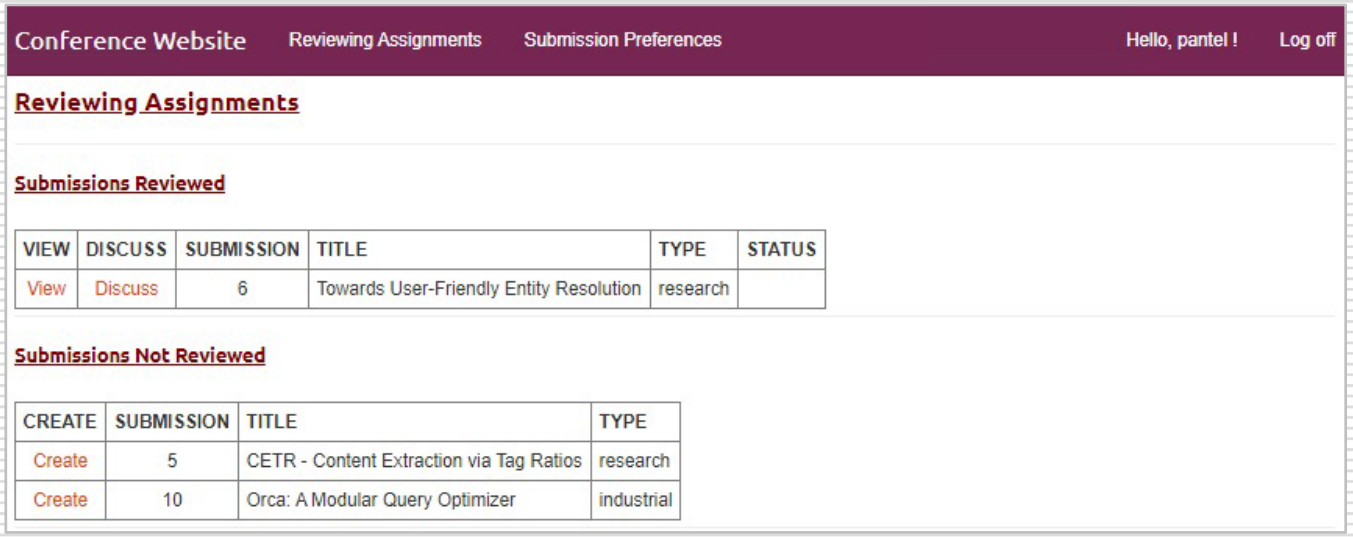


PC Chair: PC Member Information

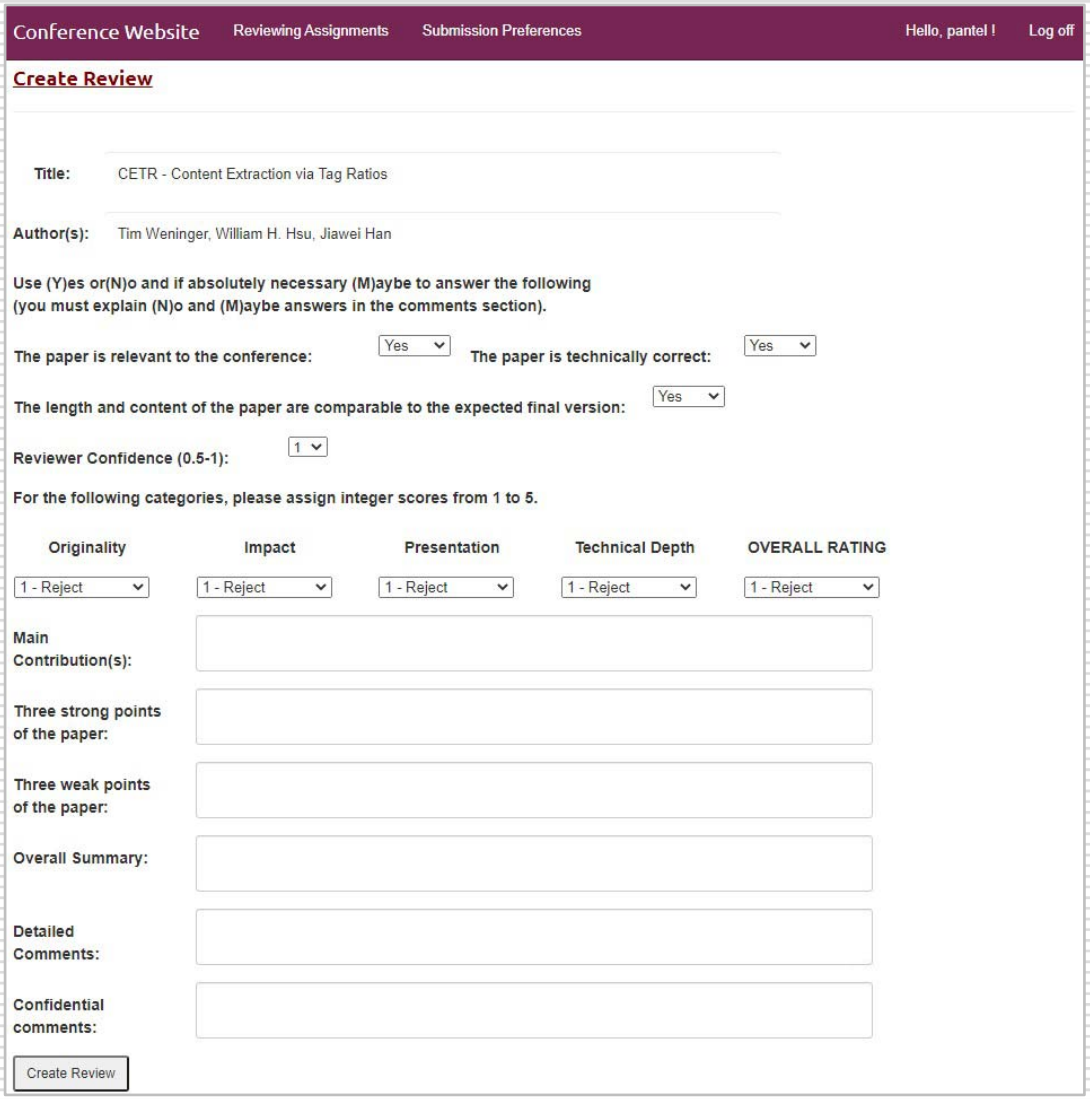


TITLE	NAME	INSTITUTION	COUNTRY	EMAIL
Prof	Beng Chin Ooi	National University of Singapore	Singapore	ooibc@comp.nus.edu.sg
Prof	Elke Rundens	Worcester Polytechnic Institute	USA	rundens@cs.wpi.edu
Dr	Eric Crestan	Yahoo! Labs	USA	ecrestan@yahoo-inc.com
Prof	Frank Lock	Hong Kong University of Science and Technology	China	lock@cse.ust.hk
Prof	H. V. Jag	University of Michigan	USA	jag@umich.edu
Prof	Jeffrey Yu	Chinese University of Hong Kong	China	yu@se.cuhk.edu.hk
Prof	Jiawei Han	University of Illinois at Urbana-Champaign	USA	hanj@cs.uiuc.edu
Prof	Lei Chen	Hong Kong University of Science and Technology	China	leichen@cse.ust.hk
Dr	Patrick Pantel	Yahoo! Labs	USA	ppantel@yahoo-inc.com
Dr	Paul Larson	Microsoft Research	USA	palarson@microsoft.com

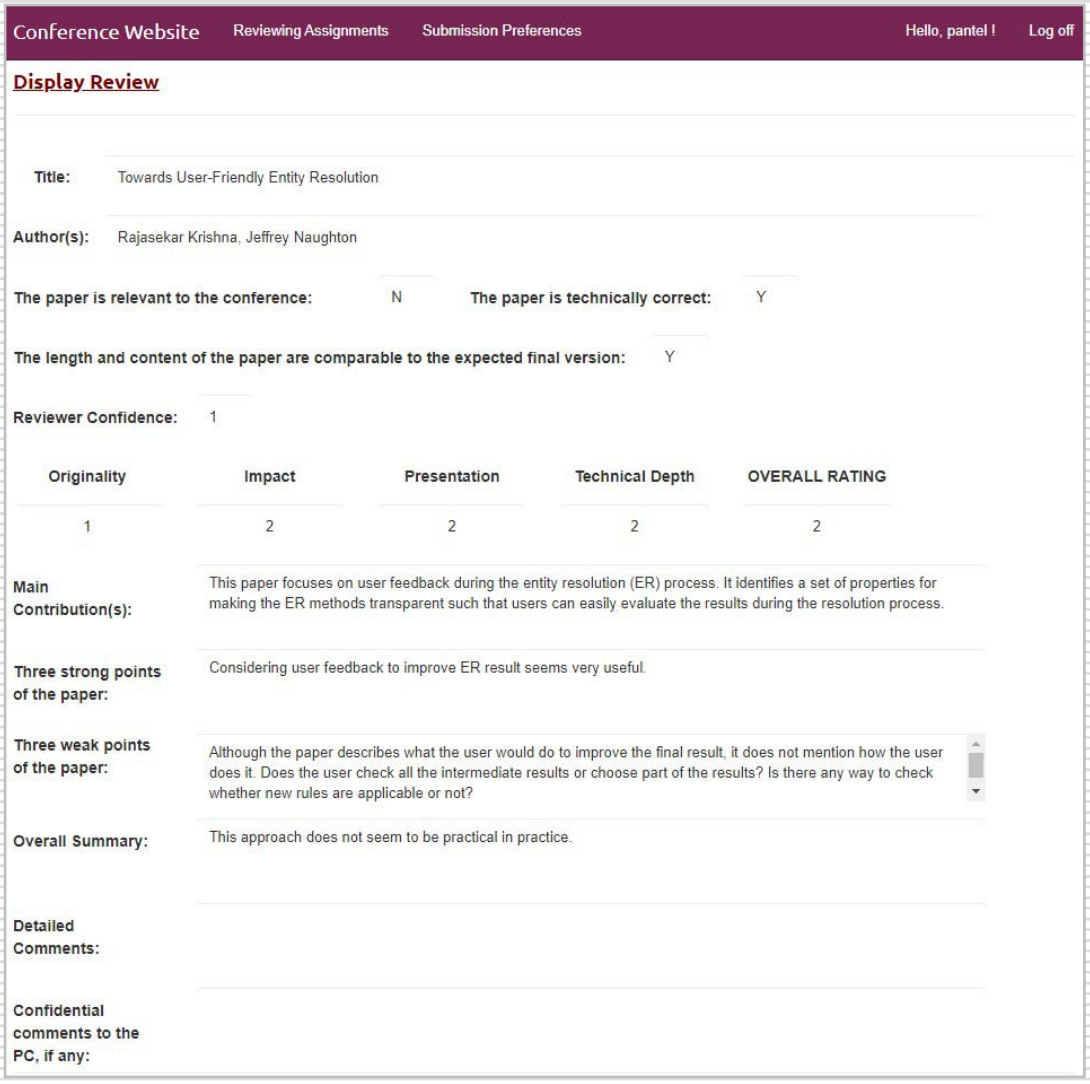
PC Member: Reviewing Assignments



PC Member: Create Review



PC Member: Display Review



Conference Website | Reviewing Assignments | Submission Preferences | Hello, pantel ! | Log off

Display Review

Title: Towards User-Friendly Entity Resolution

Author(s): Rajasekar Krishna, Jeffrey Naughton

The paper is relevant to the conference: N **The paper is technically correct:** Y

The length and content of the paper are comparable to the expected final version: Y

Reviewer Confidence: 1

Originality	Impact	Presentation	Technical Depth	OVERALL RATING
1	2	2	2	2

Main Contribution(s): This paper focuses on user feedback during the entity resolution (ER) process. It identifies a set of properties for making the ER methods transparent such that users can easily evaluate the results during the resolution process.

Three strong points of the paper: Considering user feedback to improve ER result seems very useful.

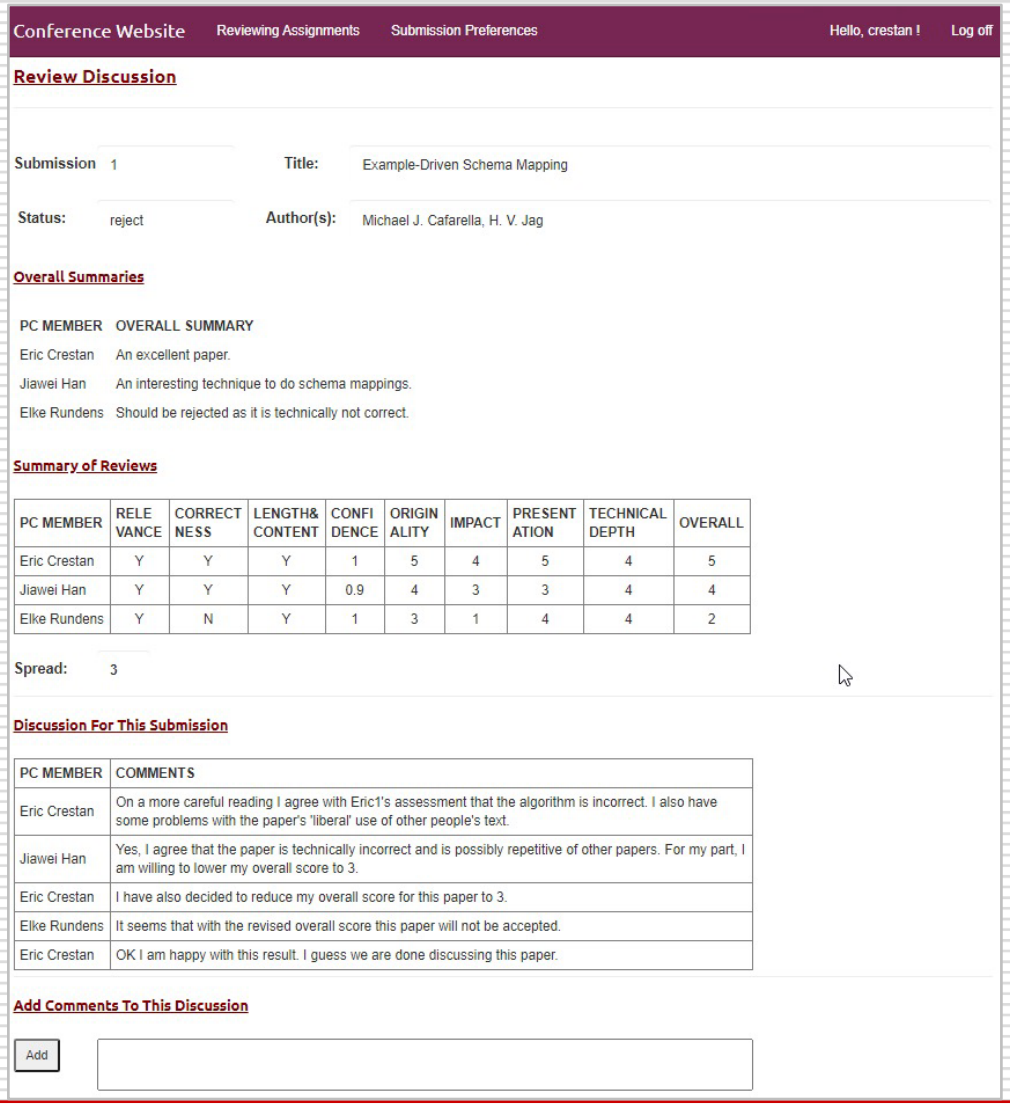
Three weak points of the paper: Although the paper describes what the user would do to improve the final result, it does not mention how the user does it. Does the user check all the intermediate results or choose part of the results? Is there any way to check whether new rules are applicable or not?

Overall Summary: This approach does not seem to be practical in practice.

Detailed Comments:

Confidential comments to the PC, if any:

PC Member: Review Discussion



Conference Website | Reviewing Assignments | Submission Preferences | Hello, crestan ! | Log off

Review Discussion

Submission 1 **Title:** Example-Driven Schema Mapping

Status: reject **Author(s):** Michael J. Cafarella, H. V. Jag

Overall Summaries

PC MEMBER	OVERALL SUMMARY
Eric Crestan	An excellent paper.
Jiawei Han	An interesting technique to do schema mappings.
Elke Rundens	Should be rejected as it is technically not correct.

Summary of Reviews

PC MEMBER	RELEVANCE	CORRECTNESS	LENGTH& CONTENT	CONFIDENCE	ORIGINALITY	IMPACT	PRESENTATION	TECHNICAL DEPTH	OVERALL
Eric Crestan	Y	Y	Y	1	5	4	5	4	5
Jiawei Han	Y	Y	Y	0.9	4	3	3	4	4
Elke Rundens	Y	N	Y	1	3	1	4	4	2

Spread: 3

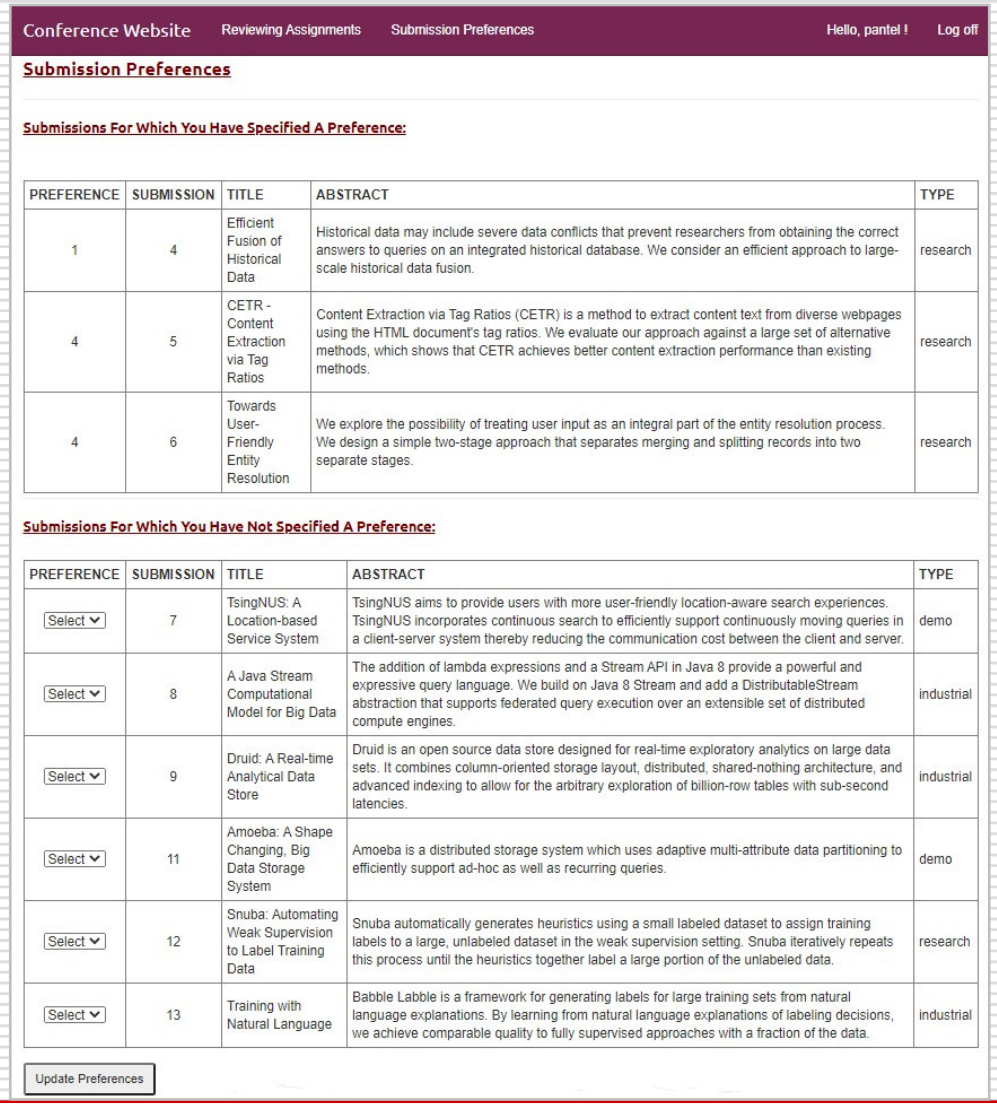
Discussion For This Submission

PC MEMBER	COMMENTS
Eric Crestan	On a more careful reading I agree with Eric1's assessment that the algorithm is incorrect. I also have some problems with the paper's 'liberal' use of other people's text.
Jiawei Han	Yes, I agree that the paper is technically incorrect and is possibly repetitive of other papers. For my part, I am willing to lower my overall score to 3.
Eric Crestan	I have also decided to reduce my overall score for this paper to 3.
Elke Rundens	It seems that with the revised overall score this paper will not be accepted.
Eric Crestan	OK I am happy with this result. I guess we are done discussing this paper.

Add Comments To This Discussion

Add

PC Member: Submission Preferences



Conference Website | Reviewing Assignments | Submission Preferences | Hello, pantel ! | Log off

Submission Preferences

Submissions For Which You Have Specified A Preference:

PREFERENCE	SUBMISSION	TITLE	ABSTRACT	TYPE
1	4	Efficient Fusion of Historical Data	Historical data may include severe data conflicts that prevent researchers from obtaining the correct answers to queries on an integrated historical database. We consider an efficient approach to large-scale historical data fusion.	research
4	5	CETR - Content Extraction via Tag Ratios	Content Extraction via Tag Ratios (CETR) is a method to extract content text from diverse webpages using the HTML document's tag ratios. We evaluate our approach against a large set of alternative methods, which shows that CETR achieves better content extraction performance than existing methods.	research
4	6	Towards User-Friendly Entity Resolution	We explore the possibility of treating user input as an integral part of the entity resolution process. We design a simple two-stage approach that separates merging and splitting records into two separate stages.	research

Submissions For Which You Have Not Specified A Preference:

PREFERENCE	SUBMISSION	TITLE	ABSTRACT	TYPE
Select	7	TsingNUS: A Location-based Service System	TsingNUS aims to provide users with more user-friendly location-aware search experiences. TsingNUS incorporates continuous search to efficiently support continuously moving queries in a client-server system thereby reducing the communication cost between the client and server.	demo
Select	8	A Java Stream Computational Model for Big Data	The addition of lambda expressions and a Stream API in Java 8 provide a powerful and expressive query language. We build on Java 8 Stream and add a DistributableStream abstraction that supports federated query execution over an extensible set of distributed compute engines.	industrial
Select	9	Druid: A Real-time Analytical Data Store	Druid is an open source data store designed for real-time exploratory analytics on large data sets. It combines column-oriented storage layout, distributed, shared-nothing architecture, and advanced indexing to allow for the arbitrary exploration of billion-row tables with sub-second latencies.	industrial
Select	11	Amoeba: A Shape Changing, Big Data Storage System	Amoeba is a distributed storage system which uses adaptive multi-attribute data partitioning to efficiently support ad-hoc as well as recurring queries.	demo
Select	12	Snuba: Automating Weak Supervision to Label Training Data	Snuba automatically generates heuristics using a small labeled dataset to assign training labels to a large, unlabeled dataset in the weak supervision setting. Snuba iteratively repeats this process until the heuristics together label a large portion of the unlabeled data.	research
Select	13	Training with Natural Language	Babble Labble is a framework for generating labels for large training sets from natural language explanations. By learning from natural language explanations of labeling decisions, we achieve comparable quality to fully supervised approaches with a fraction of the data.	industrial

Update Preferences