

# Distributed Human Computation Framework for Linked Data Co-reference Resolution

IAMRESEARCHER case study

Xiaowei Wang

ICT International Doctoral School,  
University of Trento, Italy  
xwang@disi.unitn.it

# Content

- 1. Problems in AI-computation
- 2. Available solutions
- 3. IAMRESEARCHER project
- 4. Conclusion and future work

# 1. Problem statement

Many problems are difficult for computers to solve effectively but they are relatively easy tasks for humans.

AI-Computation Fields:

- Image analysis
- Emotion recognition
- Natural language processing

# 1. Problem statement

- co-reference and linkage between datasets
- e.g.  
DBpedia <--> University of Southampton  
Researchers' information stored in both datasets

**Automatic way** of building links between them?

- Solutions?

## 2. Available solutions

Possible solutions:

- Machine learning ?
- Heuristic algorithms ?

Often turn out as following:

- Computationally expensive
- Error-prone
- Require some training data
- Difficult to deploy on a large scale

## 2. Available solutions

- **Distributed Human Computation (DHC)**

It is used to solve computational problems by incorporating the **collaborative effort of a large number of humans**.

It is also a solution to AI-computation problems

## 2. Available solutions

Co-reference solution = Social Network + DHC ??

- **Twitter/Facebook + Wikipedia**
- The combination of the Social Web and human computation may provides many opportunities to solve difficult computational problems.

## 2. Available solutions

Gruber's structure of a collective knowledge system

- users

Interacted with one another and contributed their knowledge

- machines

Facilitated communications and retrieval of resources in the background, aggregating the contributions of individual users



# 3. iamResearcher

The screenshot shows the iamResearcher website homepage. At the top, there's a navigation bar with links like 'New Tab', 'PhD Study', 'SocialLife', 'Shopping', 'Tool', 'Finance', 'Family', 'IMSC Study', and 'Career'. The main header features the 'iamResearcher BETA 0.0.8' logo and the tagline 'The Global Researcher Network'. To the right of the logo is a login section with fields for 'Email' and 'Password', a 'Remember me' checkbox, a 'Forgot password?' link, and a 'Login' button.

The central content area is divided into two main sections. On the left, a large blue box contains the text 'Get connected with researchers and resources all over the world.' Below this, it states 'iamResearcher has' followed by three statistics: '9,204,213 Publications', '20,650 Events', and '16,977 Research Institutes'. Each statistic is accompanied by a small icon. Below the statistics is a search bar with the placeholder text 'E.g. researchers name or publication title' and a 'Search' button.

On the right, a white box titled 'New to iamResearcher? Join Now. It's Free!' contains a sign-up form. The form has fields for 'First Name', 'Last Name', 'Email' (with a note 'Your email will be your login'), 'Re-enter Email', and 'Password'. A 'Sign Up' button is at the bottom of the form. Below the sign-up form, there's a link 'Create an Event (e.g. Conference, Workshop) if you are an organiser.'

The footer section contains three columns of information. The first column, titled 'What is iamResearcher?', explains the platform's purpose and includes a 'Read more >' link. The second column, titled 'For Individual Researchers', describes how to create a profile and includes a 'Read more >' link. The third column, titled 'For Conference Organisers', describes how to create an interactive conference website and includes a 'Read more >' link.

At the bottom of the page, there's a footer with links for 'About', 'Terms of Service', 'Privacy Policy', and 'DMCA Notice'. On the right, there's a social media section titled 'Follow us on Twitter (@i\_am\_researcher)' with buttons for 'Like' (124), 'Tweet' (21), and a '+1' button (1).

# 3. iamResearcher

The screenshot displays the iamResearcher web application interface. At the top, a navigation bar includes links for Help, Shortcut, Account, and a search bar. The left sidebar contains a user profile for Dr. Zhonglun Cai and a menu with options like Welcome, Homepage, Profile, Messages, Notices, Colleagues, Pages, Notes, Publications, Events, and Bookmarks. Below the menu is a calendar for December 2010. The main content area, titled 'Homepage', shows a feed of tweets and updates from Dr. Zhonglun Cai and Yang Yang. The right sidebar features a 'TODO List' with tasks, 'Recommended Publications' with links to research papers, and 'Recommended Events' with details about conferences like IEEE International Conference on Robotics and Biomimetics and ICRA 2011.

**iamResearcher** BETA 0.7

Help Shortcut Account Search Here Everything

**Dr Zhonglun Cai**  
Uni. of Southampton

Welcome  
Homepage  
Profile  
Messages 2  
Notices 4  
Colleagues

Pages  
Notes  
Publications  
Events  
Bookmarks

DECEMBER 2010

Mon	Tue	Wed	Thu	Fri	Sat	Sun
29	30	1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31	1	2
3	4	5	6	7	8	9

**Homepage**  
The latest feeds from my network.

Share: Post Bookmark Event Drop Note Talk

What's on your mind?

Colleagues Only Share

**Dr Zhonglun Cai** tweeted the following tweet.  
The Word Art of Juan Osborne | Webdesigner Depot <http://t.co/V3e7Crm>  
2 days, 3 hours ago Comment Like

**Dr Zhonglun Cai** tweeted the following tweet.  
Our Favorite Tweets of the Week Nov 22-Nov 27, 2010 | Webdesigner Depot <http://t.co/MvZ6YOH>  
2 days, 4 hours ago Comment Like

**Dr Zhonglun Cai** tweeted the following tweet.  
<http://www.hiddenpixels.com/designer-tools/brochure-design-tutorials-samples/>  
4 days, 2 hours ago Comment Like

**Yang Yang** and **Dr Zhonglun Cai** are now colleagues.  
1 week, 2 days ago Comment Like

**Dr Zhonglun Cai**  
design v7  
Download  
1 week, 3 days ago Comment Like

**Yang Yang** tweeted the following tweet.  
The YangSquared Daily is out! <http://bit.ly/9Sf80Y> ▶ Top stories today by @manusporry @jasonfried @nickbilton  
1 week, 3 days ago Comment Like

**Yang Yang** tweeted the following tweet.  
how comfortable is this guy's foot massage, well -  
<http://ww2.sinaimg.cn/bmiddle/62037b5ajw6dbk5sr55Sebj.jpg>

**TODO List**  
Write a new TODO here +  
☐ Task number 1: Finish this bloody template in time and commit the changes  
☒ Task number 2: Draw some stupid icons  
☒ Task number 3: Not sure  
☐ Task number 4: Test

**Recommended Publications**

- D. Hyndman, A. Ashburn "Stops walking when talking" as a predictor of falls in people with stroke living in the community *Journal of Neurology Neurosurgery and Psychiatry* Vol.75 pp.994 - 997 2004
- K. Galkowski, B. Sulikowski, W. Paszke, E. Rogers Stability and controllability properties of extended discrete linear repetitive processes *Proc. VII National Conference on Robotics* pp.5-14 2001
- V.N. Dubey, R.M. Crowder, P.H. Chappell, D.R. Whitley A Robotic End Effector for Unstructured Environments. *Proceedings of the American Nuclear Society's 7th Topical on Robotics and Remote Systems* pp.452-60 1997

Refresh Suggestions View All

**Recommended Events**

- IEEE International Conference on Robotics and Biomimetics**  
Location: Tianjin, China ,TIANJIN, Tianjin, China  
Time: Saturday, 18 December 2010 12:00 a.m.
- ICRA 2011 : IEEE International Conference on Robotics and Automation**  
Location: Shanghai, China ,Shanghai, Shanghai, China  
Time: Monday, 9 May 2011 12:00 a.m.
- 8th International Conference on Informatics in Control, Automation and Robotics**  
Location: Noordwijkerhout, The Netherlands  
Time: Thursday, 28 July 2011 12:00 a.m.

# 3. iamResearcher

- a social network for researchers
- running in the University of Southampton, UK
- 2008 – Now

# 3. iamResearcher

We aim to link the following two datasets

Dataset 1: papers collected from multiple repositories

Dataset 2: users registered in iamResearcher

# 3. iamResearcher

- It is really hard to make this process automatically

e.g. from Google Scholar:

- Title: Human immunoglobulin variable region gene analysis by single cell RT-PCR (biology)
- Authors: **Xiaowei Wang**, B.David Stollar,
- Title: Study on Departing Time Interval control of Bus Dispatching (business)
- Authors: Wenxia Sun; **Xiaowei Wang**; Guohui Qiao;

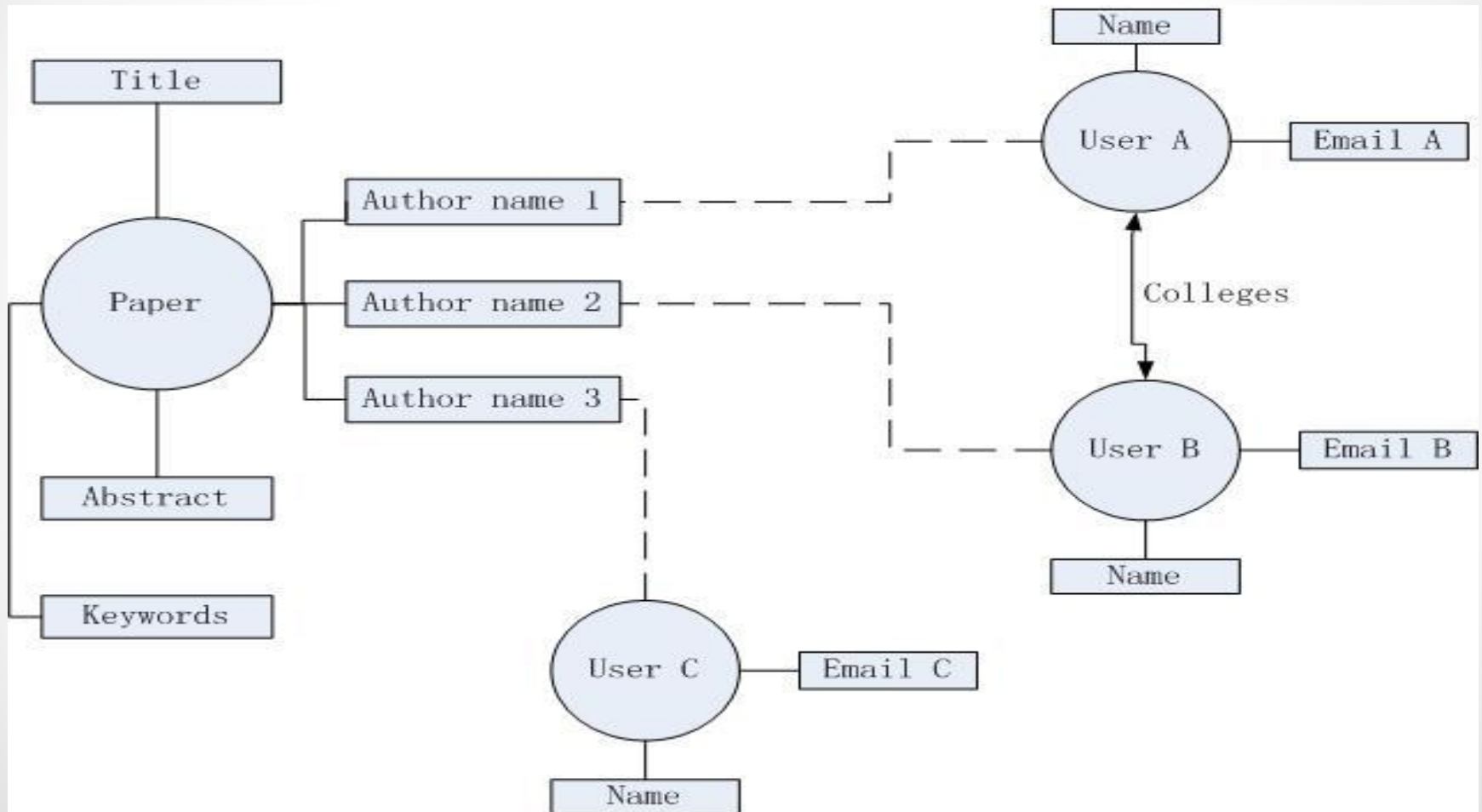
# 3. iamResearcher

- incomplete information
- information changes
- inconsistent across different digital libraries  
e.g. different styles:

X. Wang, X Wang, X. W. Wang, Wang X.,

	Full names	Names with initial
UK researcher community	10% crash	17% crash
Global researchers	?	?

# 3. iamResearcher

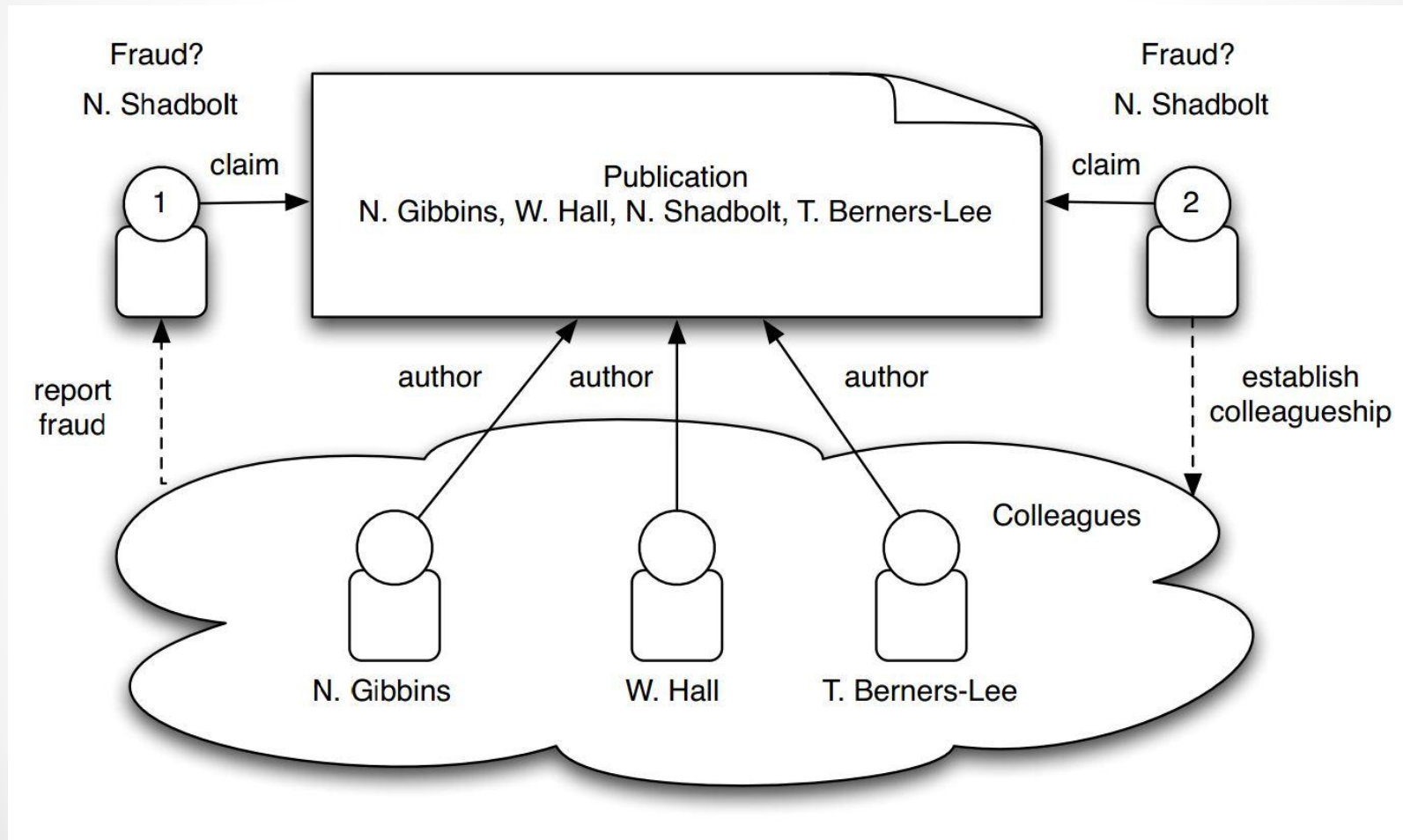


# 3. iamResearcher

<input type="checkbox"/> Collapse	Found 17 publications by the name Nick Gibbins.	①
<input checked="" type="checkbox"/> Title and Authors	Year	Claimed Users
<input checked="" type="checkbox"/> Effective Benchmarking for RDF Stores Using Synthetic Data A Owens , Alisdair Owens , Nick Gibbins , mc schraefel	2008	
<input checked="" type="checkbox"/> Transitioning Applications to Semantic Web Services: An Automated Formal Approach Hai H. Wang , Nick Gibbins , Terry Payne , Ahmed Saleh , Yuan Li	2008	
<input checked="" type="checkbox"/> Formal Specification of OWL-S with Object-Z: the Static Aspect H.H. Wang , Hai H. Wang , Ahmed Saleh , Terry Payne , Nick Gibbins	2007	
<input checked="" type="checkbox"/> An Integrated Formal Approach to Semantic Work Environments Design Hai H. Wang , Jin-Song Dong , Jing Sun , Terry R. Payne , Nick Gibbins , Yuan-Fang Li , Jeff Pan	2007	
<input checked="" type="checkbox"/> Connecting Physical+Temporal Events to Digital Contexts monica m.c. schraefel , Jan Millard , Dave Millard , Gareth V. Hughes , Mark J. Weal , Danius T. Michaelides , Steve Harris , Nick Gibbins	2003	
<input type="checkbox"/> Collapse	Found 31 publications by the name N Gibbins.	②
<input type="checkbox"/> Title and Authors	Year	Claimed Users
<input type="checkbox"/> D4. 3.1 Annex: Distributed Indexing and Querying of Distributed Knowledge N Gibbins , A Douglas		Nicholas Gibbins
<input type="checkbox"/> Resource Description Framework (RDF) N Gibbins , N Shadbolt	2009	Nicholas Gibbins
<input type="checkbox"/> User-Induced Hyperlinks in Collaborative Tagging Systems CA Yeung , N Gibbins , N Shadbolt	2009	Nicholas Gibbins
<input type="checkbox"/> Studies on Editing Patterns in Large-scale Wikis P Boulain , N Shadbolt , N Gibbins	2009	Nicholas Gibbins
<input type="checkbox"/> Cs aktivespace: Building a semantic web application H. Glaser , H. Alani , L. Carr , S. Chapman , F. Ciravegna , A. Dingli , N. Gibbins , S. Harris , M. Schraefel , N. Shadbolt	2004	Nicholas Gibbins
<input type="checkbox"/> Put in your postcode, out comes the data: A Case Study T. Omitola , C.L. Koumenides , I.O. Popov , Y. Yang , M. Salvadores , M. Szomszor , T. Berners-Lee , N. Gibbins , W. Hall , N. Shadbolt	2010	Nicholas Gibbins
<input type="checkbox"/> Expand	Found 1 publication by the name N.M. Gibbins	③
<input type="checkbox"/> Expand	Found 1 publication by the name NM Gibbins.	④



# 3. iamResearcher



# 3. iamResearcher



## **Showcase your work**

- Create your profile with your publications and research achievements so that other researchers and organisations can see your works.



## **Link with your colleagues**

- Network with researchers, follow experts in your research field and share the latest research news and publications.



## **Stay updated**

- iamResearcher tells you about upcoming research workshops, conferences and funding bids. All directly related to your research interests.

# 3. iamResearcher

Total Users	Users have published	Users claimed	Percentage
163	52	39	75%

Preliminary data showed that people did use the function

Date taken by 2011.02,  
the web site was just published 2 months in University of Southampton

# 4. Conclusion & future work

- DHCF facilitates some AI-computation difficulties

iamResearcher:

- Link authors to their papers
- The authors are connected in a social network
- Tasks are distributed to the users in the system
- Form an ecosystem:

self-sufficient system that can provide a long-term solution to a particular problem

# 4. Conclusion & future work

Merge AI with DHCF

- Use email, co-author, keywords information to facilitate users' work of identifying their papers

Track the editing history

- Create a log system like wikipedia

Enhance the functions for researchers

- Publication, Group, Schedule, search engine

# The end

# Thanks!