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

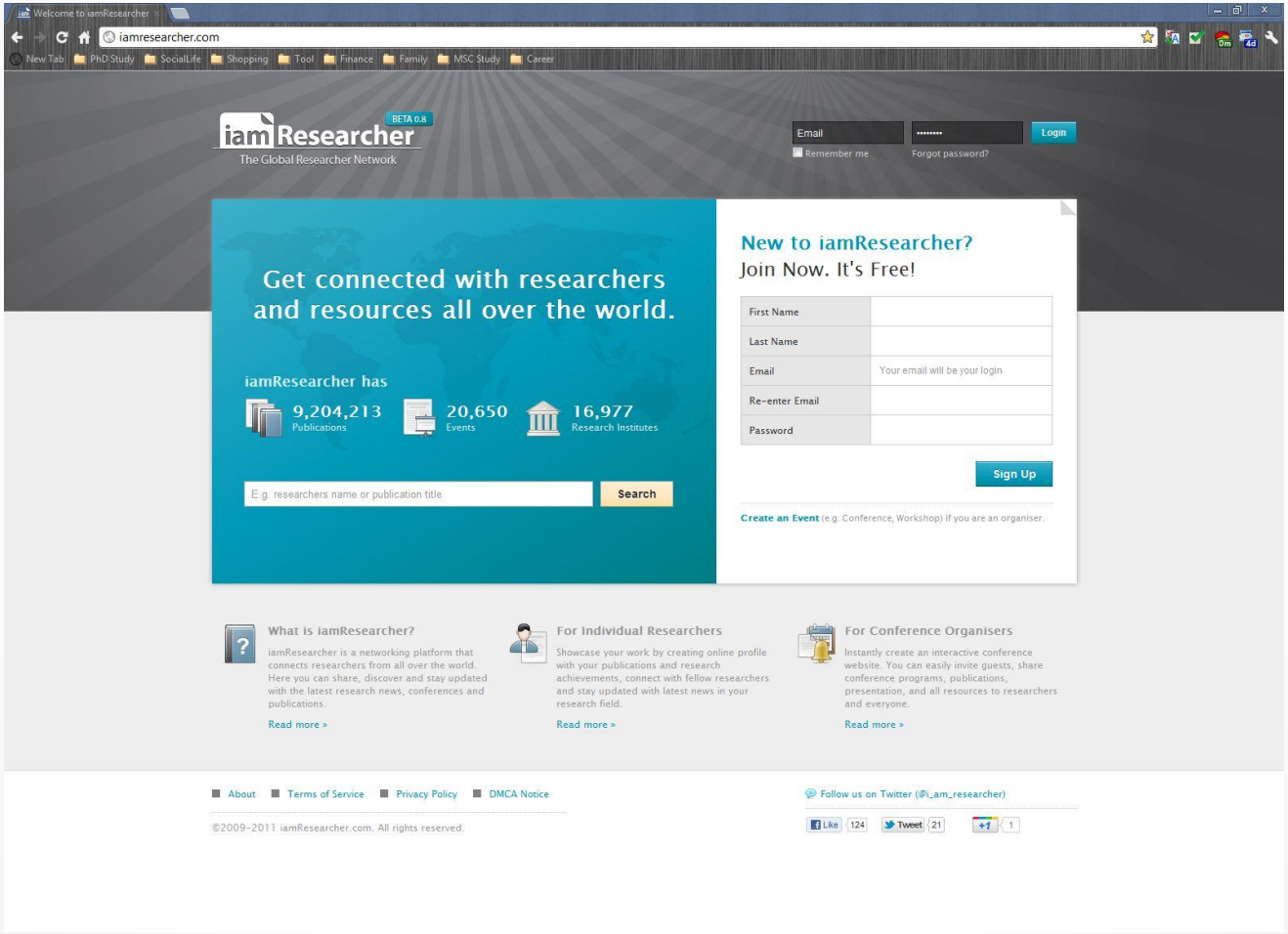
- users

Interacted with one another and contributed their knowledge

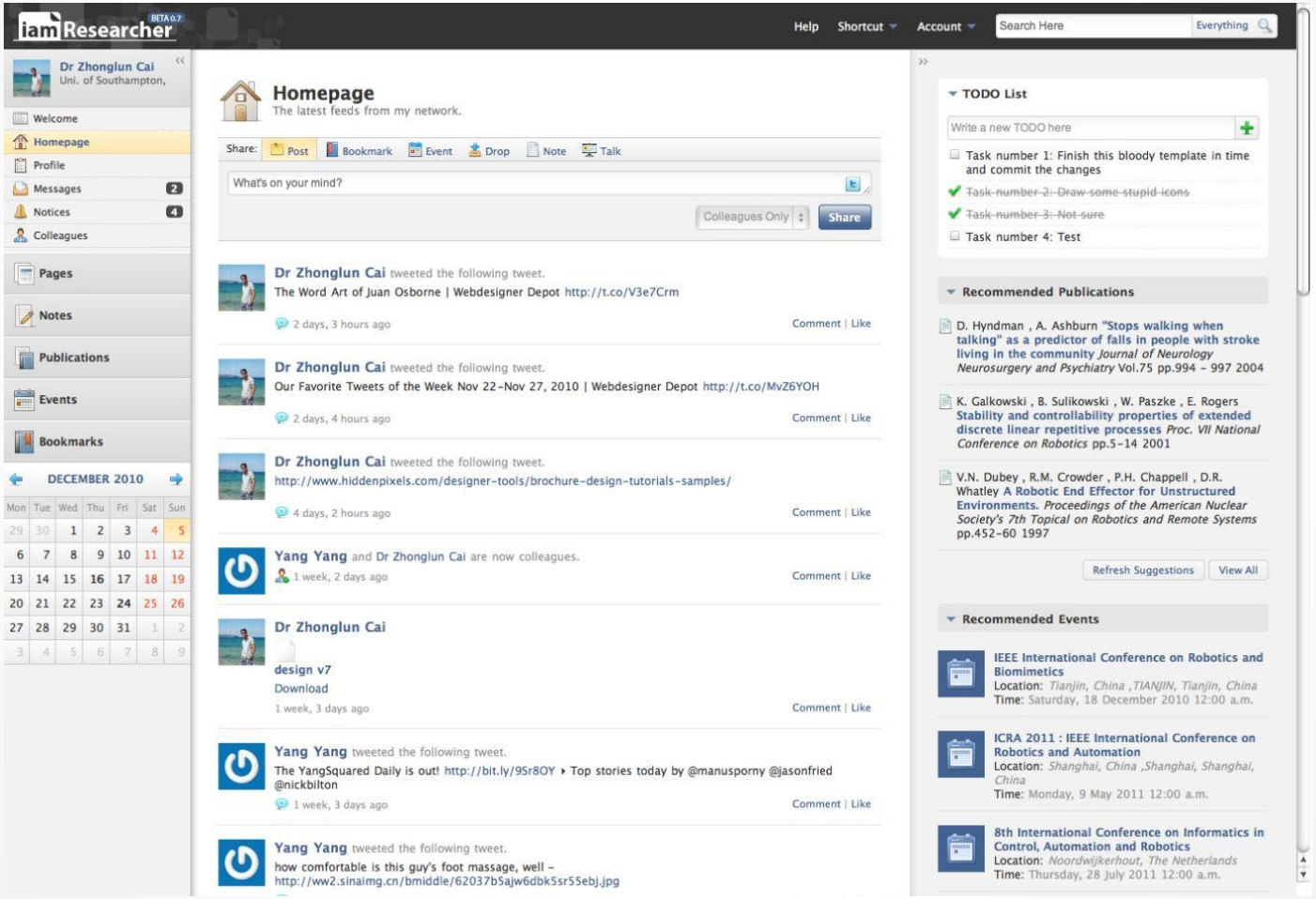
- machines

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

3. iamResearcher



3. iamResearcher



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

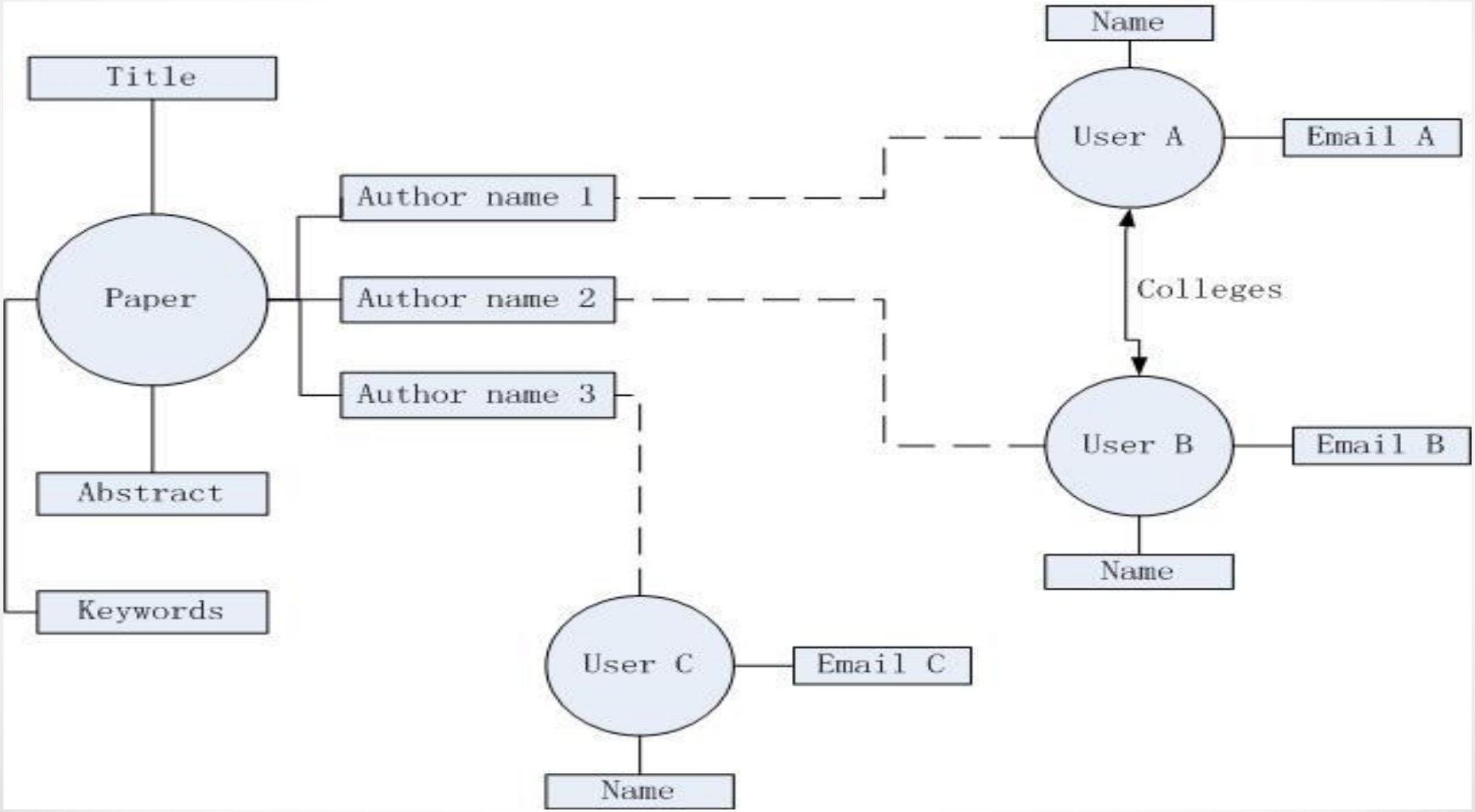
- incomplete information
- information changes
- inconsistent across different digital libraries

  e.g. different styles:

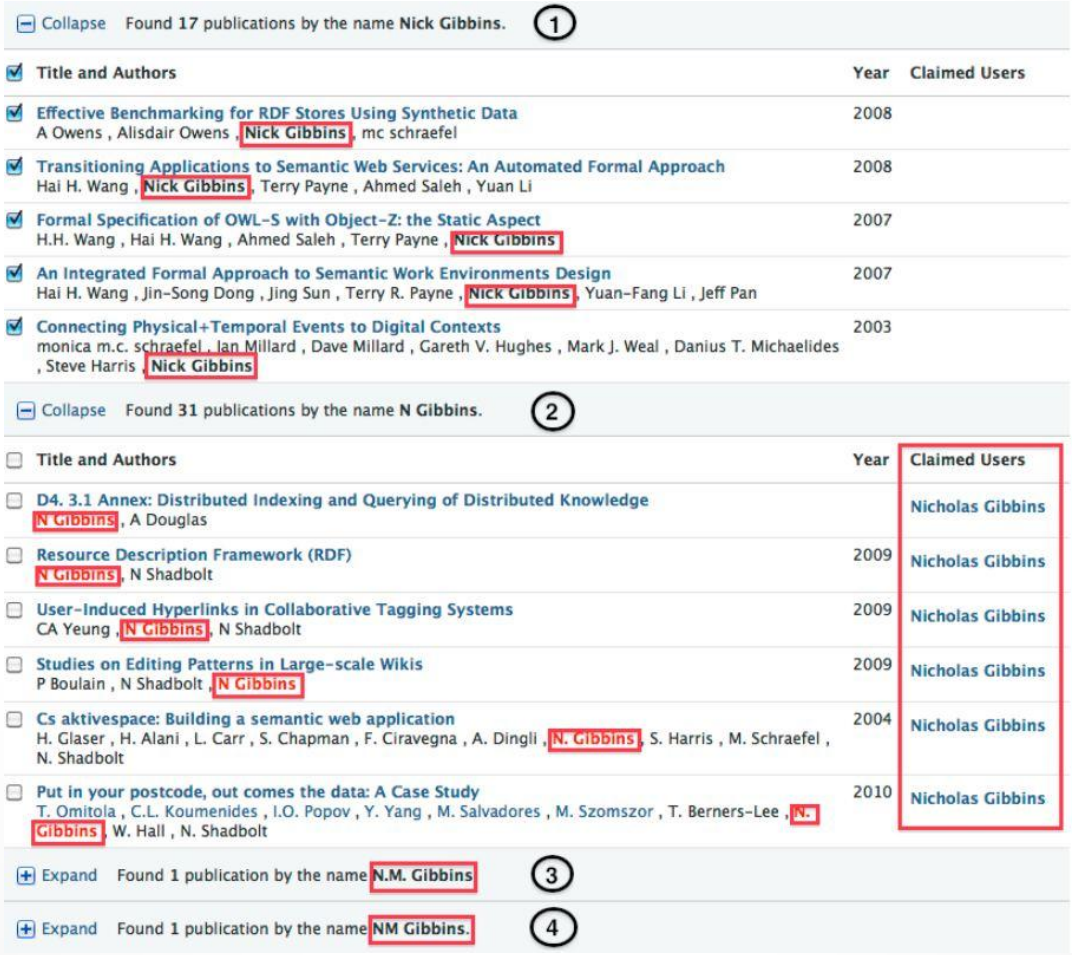
  X. Wang, X Wang, X. W. Wang, Wang X.,

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

3. iamResearcher



3. iamResearcher



Collapse Found 17 publications by the name Nick Gibbins. (1)

Title and Authors	Year	Claimed Users
Effective Benchmarking for RDF Stores Using Synthetic Data A Owens , Alisdair Owens , Nick Gibbins , mc schraefel	2008	
Transitioning Applications to Semantic Web Services: An Automated Formal Approach Hai H. Wang , Nick Gibbins , Terry Payne , Ahmed Saleh , Yuan Li	2008	
Formal Specification of OWL-S with Object-Z: the Static Aspect H.H. Wang , Hai H. Wang , Ahmed Saleh , Terry Payne , Nick Gibbins	2007	
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	
Connecting Physical+Temporal Events to Digital Contexts monica m.c. schraefel , Ian Millard , Dave Millard , Gareth V. Hughes , Mark J. Weal , Danius T. Michaelides , Steve Harris , Nick Gibbins	2003	

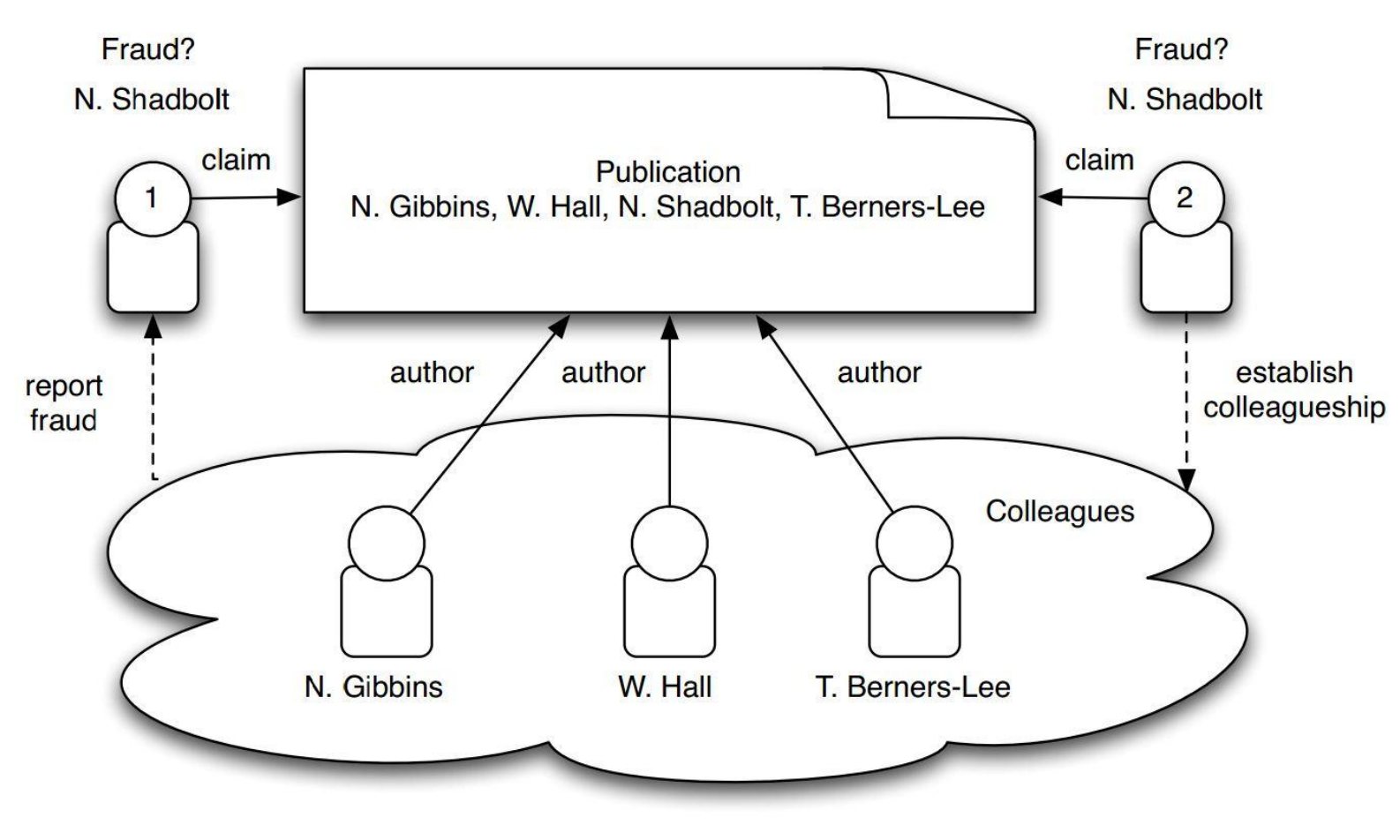
Collapse Found 31 publications by the name N Gibbins. (2)

Title and Authors	Year	Claimed Users
D4. 3.1 Annex: Distributed Indexing and Querying of Distributed Knowledge N Gibbins , A Douglas		Nicholas Gibbins
Resource Description Framework (RDF) N Gibbins , N Shadbolt	2009	Nicholas Gibbins
User-Induced Hyperlinks in Collaborative Tagging Systems CA Yeung , N Gibbins , N Shadbolt	2009	Nicholas Gibbins
Studies on Editing Patterns in Large-scale Wikis P Boulain , N Shadbolt , N Gibbins	2009	Nicholas Gibbins
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
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

Expand Found 1 publication by the name N.M. Gibbins (3)

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 identifing 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!