
Supporting Online Problem-Solving Communities with the Semantic Web

Anupriya Ankolekar

University of Karlsruhe

Katia Sycara

James Herbsleb

Robert Kraut

Carnegie Mellon University

Chris Welty

IBM Watson Research

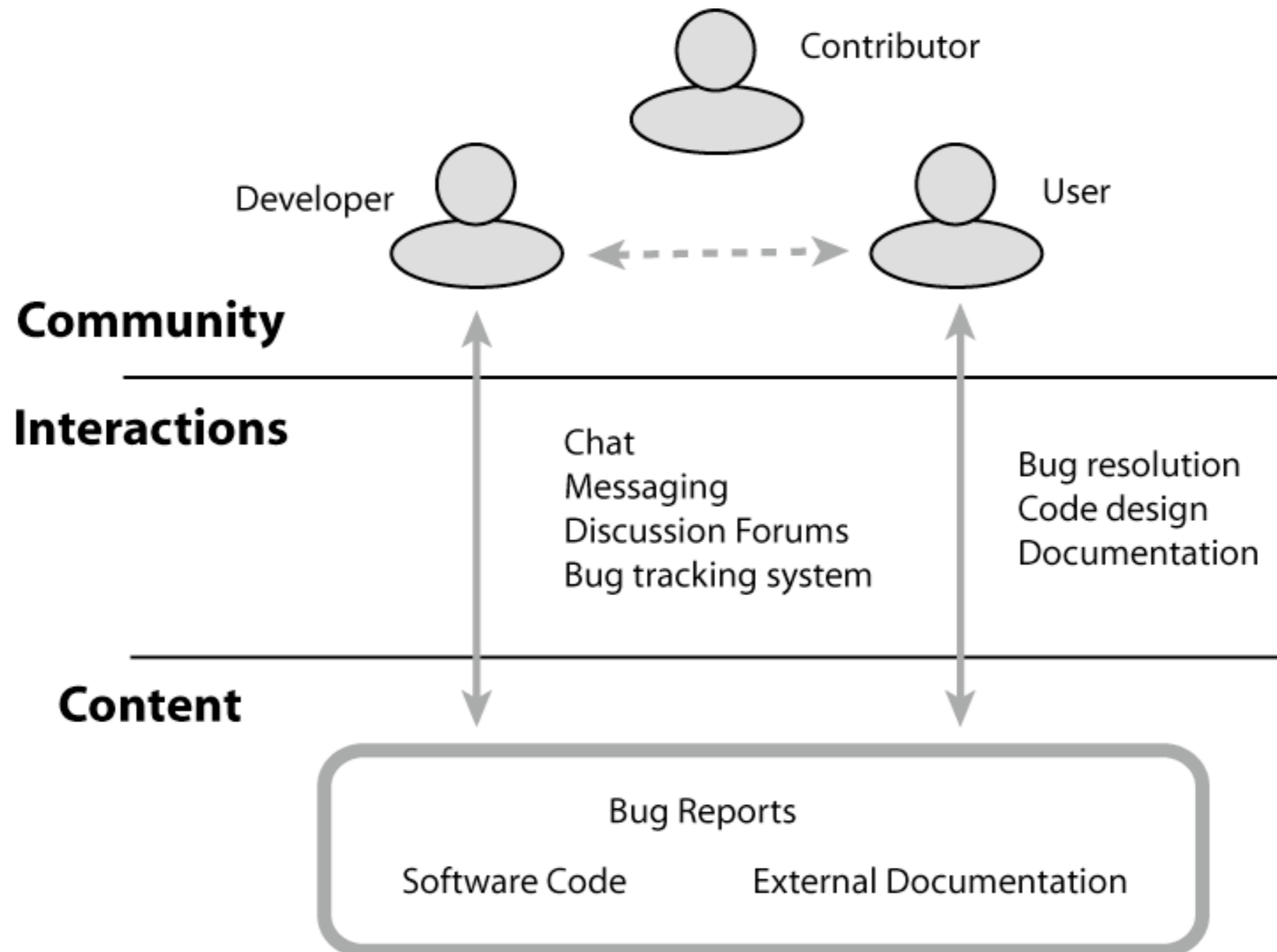
Problem-solving communities on the Web

- ▶ open source software (OSS) communities
- ▶ engineering design communities
- ▶ e-learning communities
- ▶ medical research communities

- ▶ Professionals using the Web to collaboratively
 - diagnose problems
 - design and create artefacts
 - analyse and understand information



Open Source Software (OSS) Communities



A typical OSS Community ...



OpenACS/dotLRN

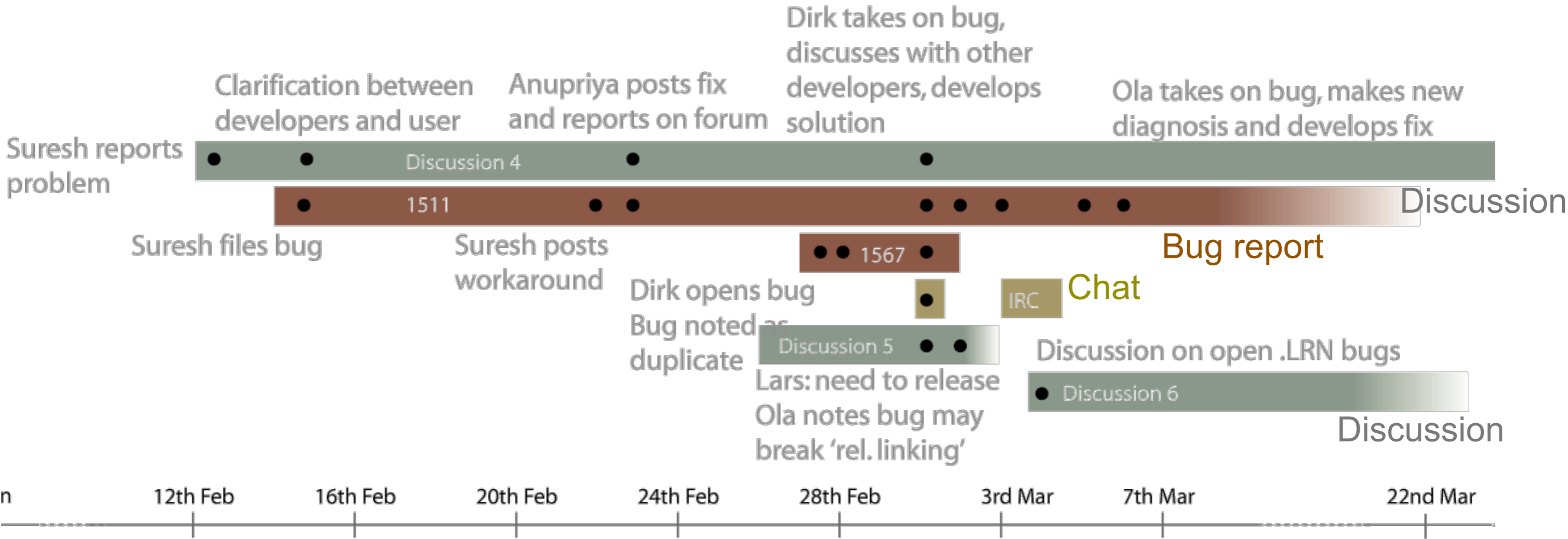
- Toolkit for building Web community-oriented applications: mailing lists, forums, shared files and calendars, chat, issue tracking system
- Extended to e-learning platform dotLRN (used by MIT, Uni. Heidelberg, ...)
- Community interactions through in-house tools

The screenshot shows the OpenACS website interface. At the top, there is a navigation bar with links for News, Forums, Community, Projects, Bugs, Documentation, and Download. Below this, a header section includes the OpenACS logo, the tagline 'The Toolkit for Online Communities', and statistics: '9880 Community Members, 1 member online'. There are also links for 'Log In' and 'Register', and a search box.

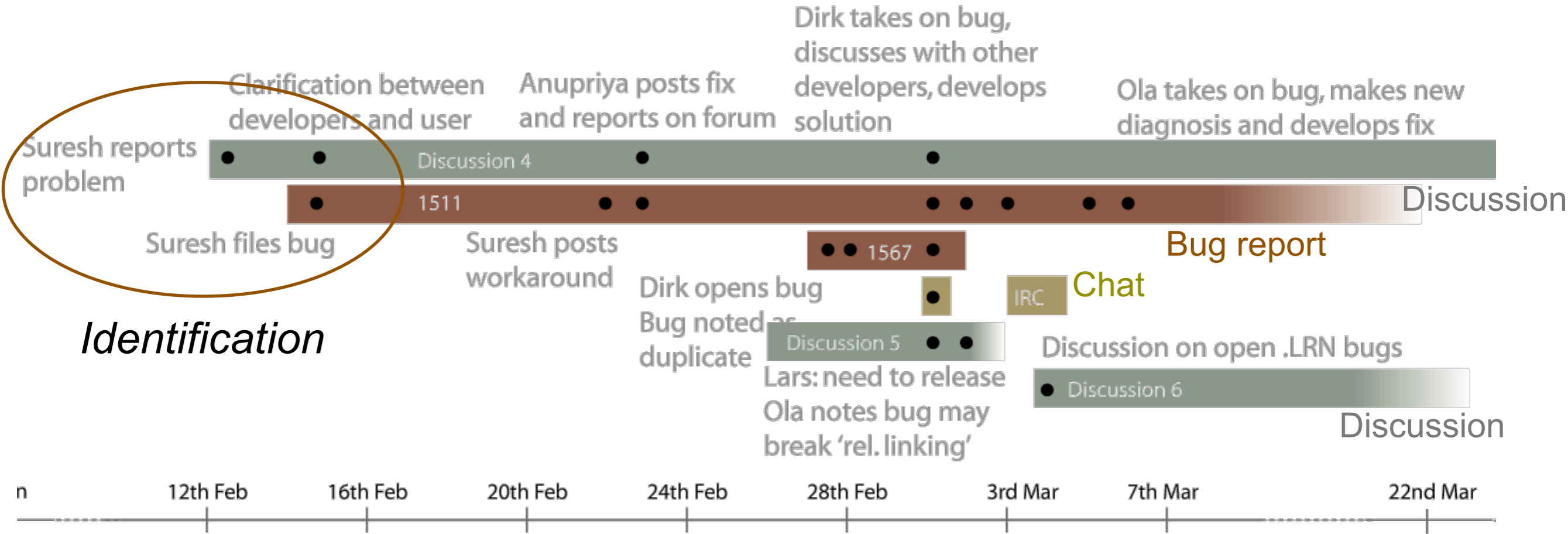
The main content area is divided into several sections:

- News:** A list of recent news items with dates and titles, such as 'May 14, 2006 Spring 2006 OCT Election - Voting Begins' and 'May 03, 2006 OpenACS Core 5.2.3 Release Candidate 1 released'. There is also a 'Syndication Feed' link with an XML icon.
- Recent Posts:** A list of forum posts, including '.LRN Q&A another annoying spam post', 'OpenACS CMS ActionApps to OpenACS (MySQL to PostgreSQL)', and 'OpenACS Development AMS package ams_options'.
- Main Content:** A central article titled 'OpenACS (Open Architecture Community System) is a toolkit for building scalable, community-oriented web applications...'. It includes sections for 'Why use OpenACS?', 'How do I install OpenACS?', and 'How do I work with OpenACS?'. The text describes OpenACS as an open-source toolkit for building web applications, mentioning its use in products like dotLRN and its availability under the GNU General Public License.
- Right Sidebar:** Contains several informational boxes: 'Unleash OpenACS NOW!', 'About OpenACS' (with links for 'Why use OpenACS?', 'Screenshots', 'Project status', 'History of OpenACS', 'Learning OpenACS', 'Licensing', 'OpenACS.org Web Site Credits', 'OpenACS Team', 'FAQs', and 'Contribute to OpenACS'), and 'Features Articles' (with a link for 'OpenACS, Robust web development framework').

Identification and resolution of a bug

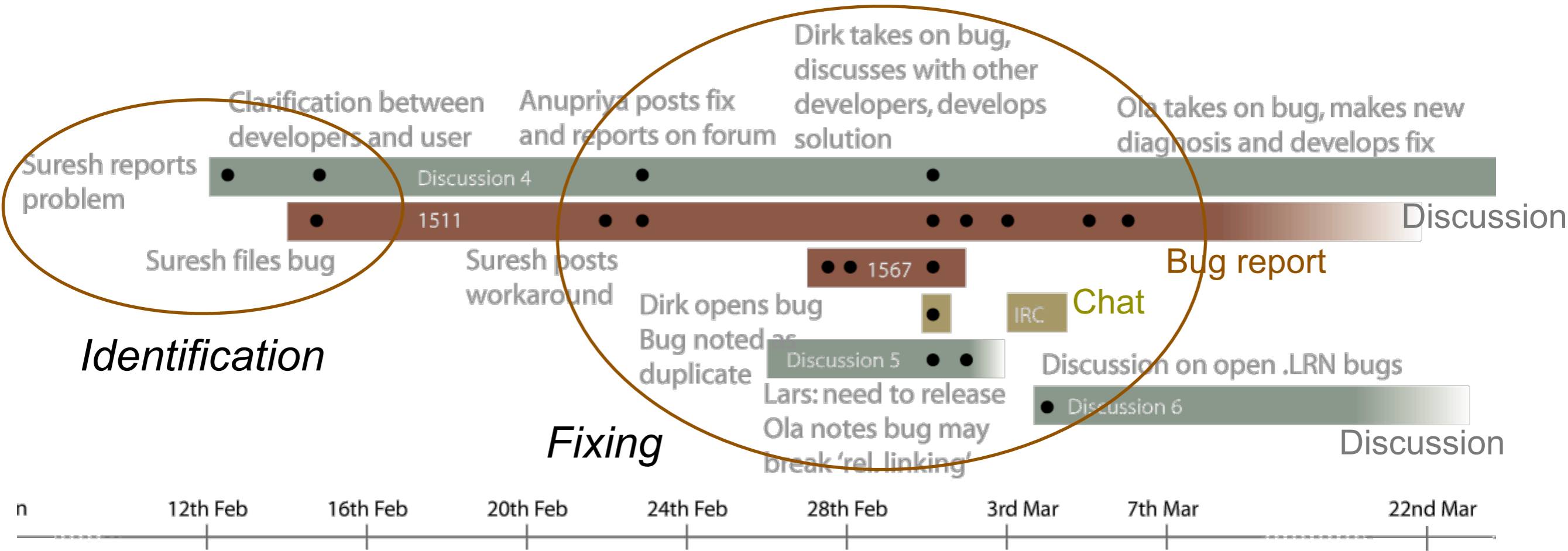


Identification and resolution of a bug

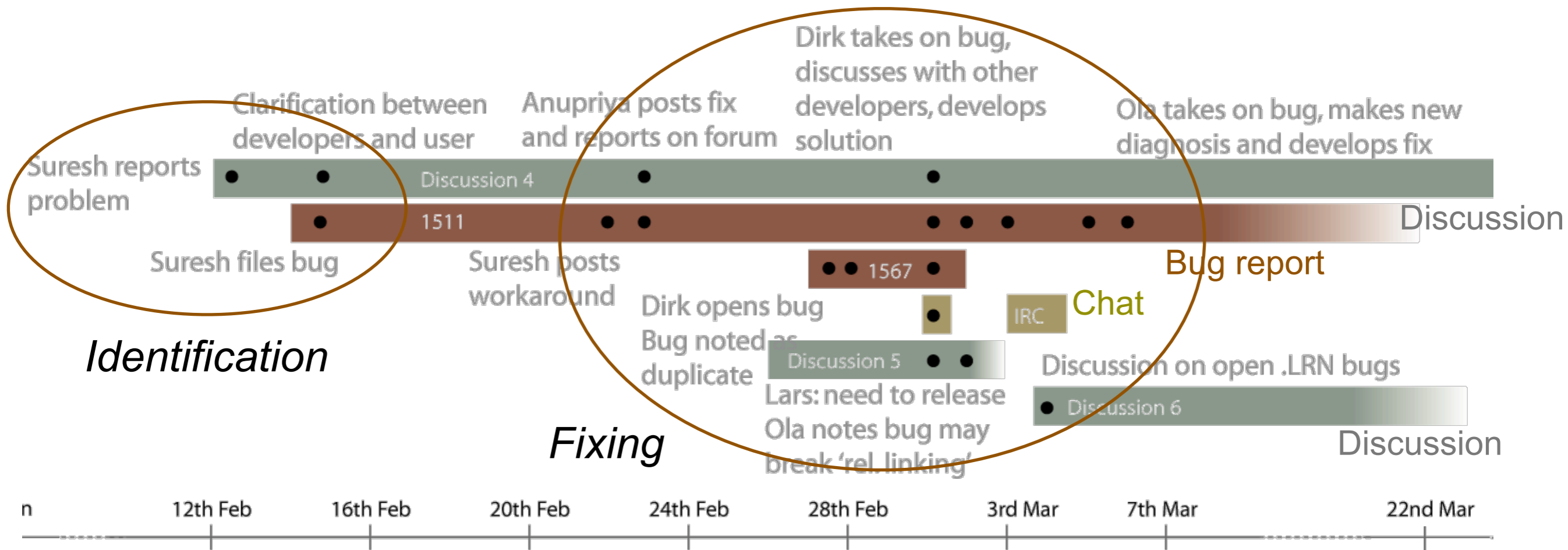


Identification

Identification and resolution of a bug

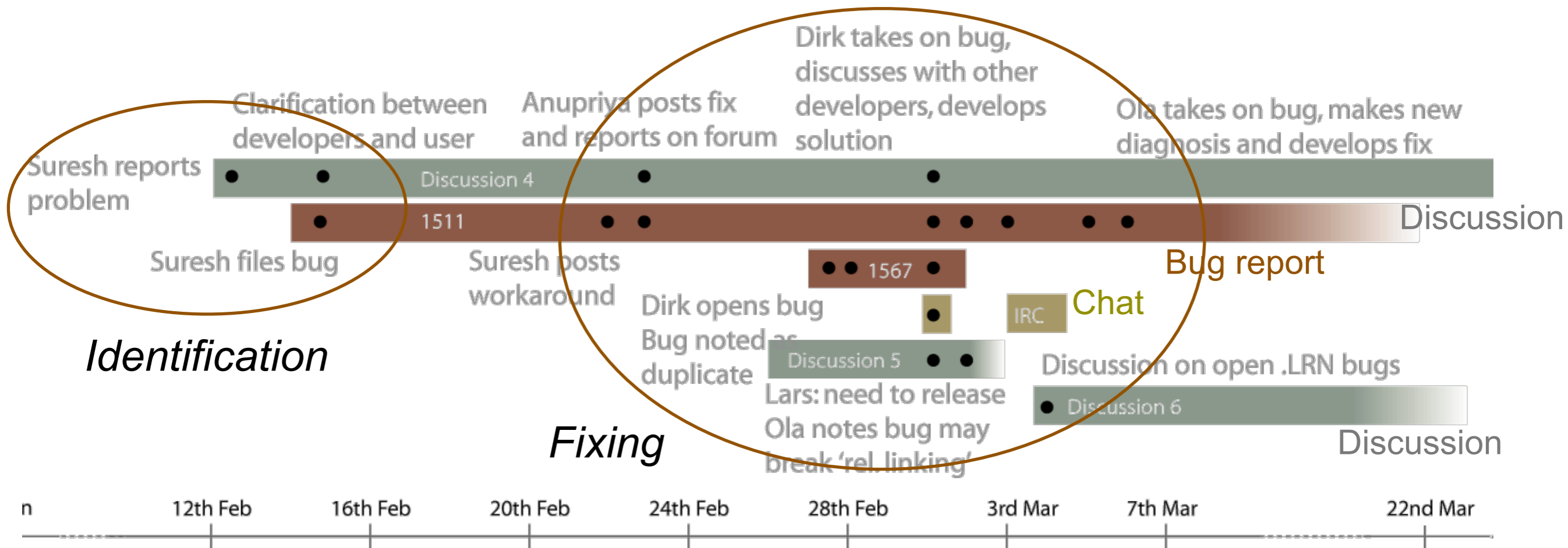


Identification and resolution of a bug



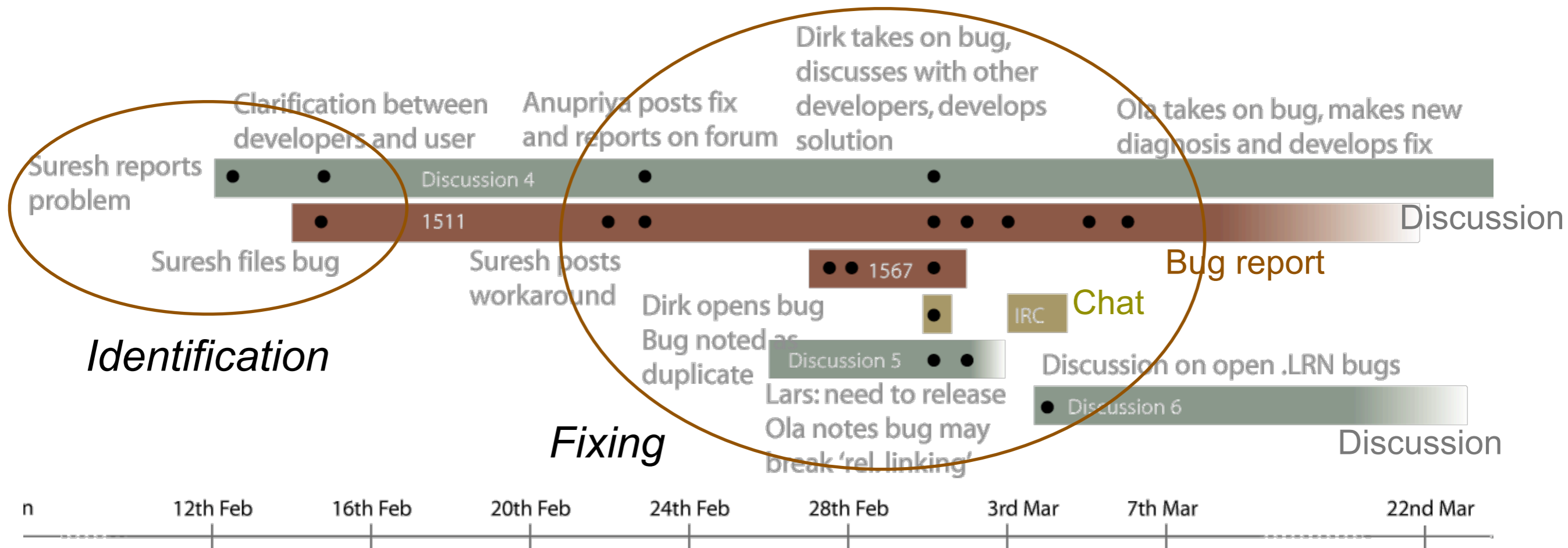
- ▶ Discussion about bug over several channels

Identification and resolution of a bug



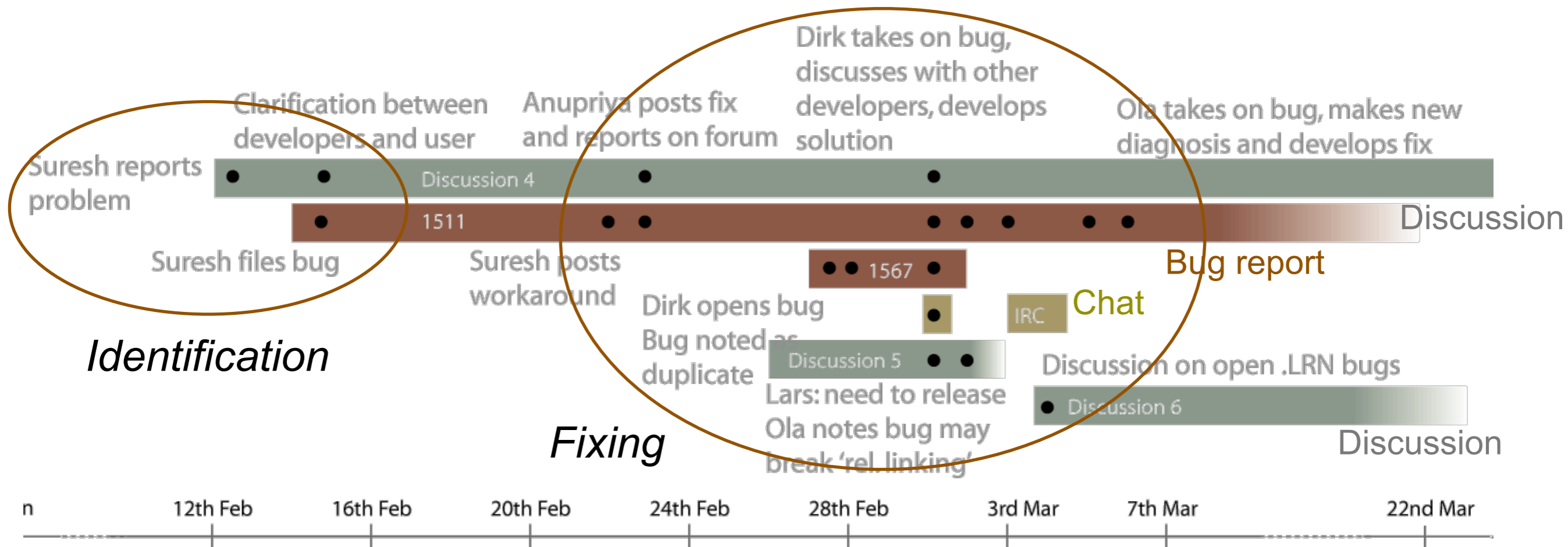
- ▶ Discussion about bug over several channels
- ▶ Duplicate bug filed

Identification and resolution of a bug



- ▶ Discussion about bug over several channels
- ▶ Duplicate bug filed
- ▶ Poorly linked information

Identification and resolution of a bug



- ▶ Discussion about bug over several channels
- ▶ Duplicate bug filed
- ▶ **Poorly linked information**
- ▶ More than one month to fix the bug

How to interpret a bug report message?

1/02/2004 Opened by Tilmann Singer

forums-security-procs.tcl was changed to check for standard privs instead of the packages custom privs, e.g. create instead of forums_create etc., but

1) The privileges are still created upon package installation. This is very confusing for users since as far as I can see, apart from forums_moderate they have no effect anymore.

2) They are still used, at least in forum::posting_policy_set and forum::new_questions_allow, which from reading the code I assume don't work anymore, since they grant for example 'forum_create' but forums-security-procs.tcl require 'create' now.

How to interpret a bug report message?

1/02/2004 Opened by Tilmann Singer

`forums-security-procs.tcl` was changed to check for `standard privs` instead of the `packages custom privs`, e.g. `create` instead of `forums_create` etc., but

1) The `privileges` are still created upon package installation. This is very confusing for users since as far as I can see, apart from `forums_moderate` they have no effect anymore.

2) They are still used, at least in `forum::posting_policy_set` and `forum::new_questions_allow`, which from reading the code I assume don't work anymore, since they grant for example `'forum_create'` but `forums-security-procs.tcl` require `'create'` now.

How to interpret a bug report message?

File: forums-security-procs.tcl

Bug: Ability to close a thread

File: forums-procs.tcl

Doc: OpenACS Permissions Tediously Explained

File: forums-security-procs.tcl

1/02/2004 Opened by Tilmann Singer

forums-security-procs.tcl was changed to check for standard privs instead of the packages custom privs, e.g. create instead of forums_create etc., but

1) The privileges are still created upon package installation. This is very confusing for users since as far as I can see, apart from forums_moderate they have no effect anymore.

2) They are still used, at least in forum::posting_policy_set and forum::new_questions_allow, which from reading the code I assume don't work anymore, since they grant for example 'forum_create' but forums-security-procs.tcl require 'create' now.

Person: Lars Pind

Person: Jeff Davis

Disc: Custom privileges for a package

Disc: Best Practices for permissions

Disc: Howto: Expand and Use acs-subsite Based Member Roles

How to interpret a bug report message?

File: forums-security-procs.tcl

Bug: Ability to close a thread

File: forums-procs.tcl

Doc: OpenACS Permissions Tediously Explained

File: forums-security-procs.tcl

1/02/2004 Opened by Tilmann Singer

forums-security-procs.tcl was changed to check for standard privs instead of the packages custom privs, e.g. create instead of forums_create etc., but

1) The privileges are still created upon package installation. This is very confusing for users since as far as I can see, apart from forums_moderate they have no effect anymore.

2) They are still used, at least in forum::posting_policy_set and forum::new_questions_allow, which from reading the code I assume don't work anymore, since they grant for example 'forum_create' but forums-security-procs.tcl require 'create' now.

Person: Lars Pind

Person: Jeff Davis

Disc: Custom privileges for a package

Disc: Best Practices for permissions

Disc: Howto: Expand and Use acs-subsite Based Member Roles

Interpreting content of message requires knowledge of technical terms, source code structure and past discussions

Supporting bug resolution

Two distinguishing characteristics of problem-solving in bug resolution:

1. Knowledge of the past history of interactions about a given topic is crucial
2. Exploration of relations between code objects

Locating files, discussions and bug reports, and finding experts is not straightforward search problem:

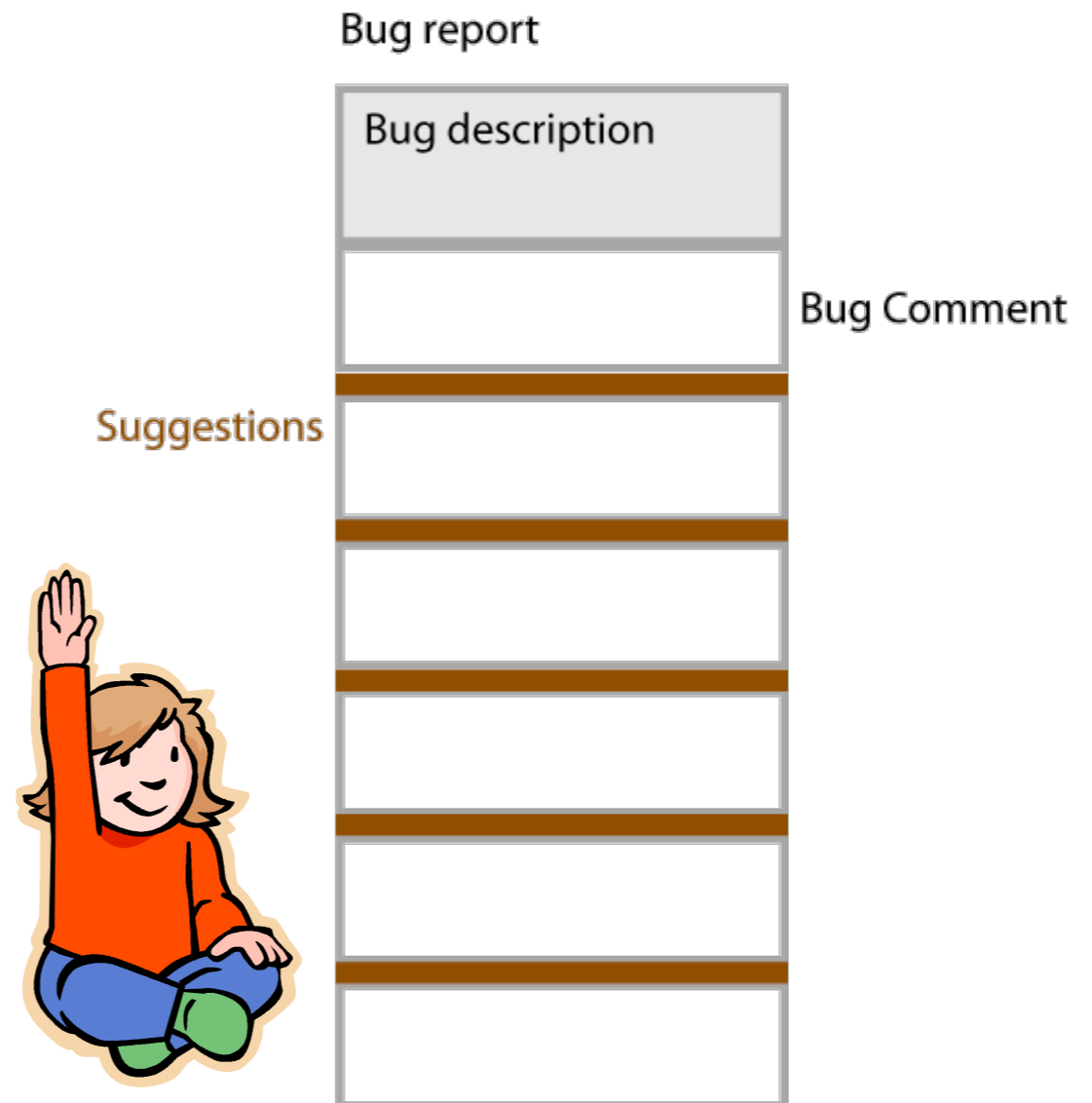
- What are the keywords to search for?
- Not obvious which artifacts are going to be helpful finally
- Person who wrote the code may not be around or person who discussed problem may not know its implementation details

Semantic Web system to support bug resolution

... monitor bug report messages

For each bug report message

- ▶ Identify 'semantics' of message
- ▶ Use 'semantics' to search for similar resources
 - Discussions
 - Bug reports
 - Files
 - People
 - ...
- ▶ Use heuristics to recommend relevant resources



What does the Semantic Web provide?



3/01/2004 Resolved (Fixed) by [Dirk Gomez](#)

The URL for a file-storage folder contains I18Nized strings and the tcl pages don't localize them. Furthermore some of the links were missing `@?version_id=@contents.live_revision@`. Fixed that.

- Metadata
 - Links: references to code, files, packages, error traces, other bug reports and discussions
 - Semantic concepts: technical vocabulary meaningful to community members
- Ontologies
 - Content: Where are links and semantic concepts in software code and bugs?
 - To identify what the bug report message is about
 - Interactions: Who has done what when?
 - To identify people who are experts in area of bug
 - Community: Who typically does what?
 - To make suggestions appropriate to roles of experts

What does the Semantic Web provide?



3/01/2004 Resolved (Fixed) by [Dirk Gomez](#)

The URL for a `file-storage` folder contains I18Nized strings and the tcl pages don't localize them. Furthermore some of the links were missing `@?version_id=@contents.live_revision@`. Fixed that.

- Metadata
 - Links: references to code, files, packages, error traces, other bug reports and discussions
 - Semantic concepts: technical vocabulary meaningful to community members
- Ontologies
 - Content: Where are links and semantic concepts in software code and bugs?
 - To identify what the bug report message is about
 - Interactions: Who has done what when?
 - To identify people who are experts in area of bug
 - Community: Who typically does what?
 - To make suggestions appropriate to roles of experts

What does the Semantic Web provide?



3/01/2004 Resolved (Fixed) by [Dirk Gomez](#)

The URL for a file-storage folder contains I18Nized strings and the tcl pages don't localize them. Furthermore some of the links were missing `@?version_id=@contents.live_revision@`. Fixed that.

- Metadata
 - Links: references to code, files, packages, error traces, other bug reports and discussions
 - Semantic concepts: technical vocabulary meaningful to community members
- Ontologies
 - Content: Where are links and semantic concepts in software code and bugs?
 - To identify what the bug report message is about
 - Interactions: Who has done what when?
 - To identify people who are experts in area of bug
 - Community: Who typically does what?
 - To make suggestions appropriate to roles of experts

Semantic Web Reserach

Research focus on tools and techniques

- ▶ Ontologies and annotation
[Borgida, Brachman 03] [Welty 02] [Handschuh et al. 03] ...
- ▶ Reasoning
[Horrocks, Patel-Schneider, van Harmelen 03] [Haarslev, Möller 01] [Kopena, Regli 03] ...
- ▶ Applications
 - Semantic Web services [Sycara et al. 03] ...
 - Semantic Portals [Maedche et al. 03]
 - E-learning [Schmitz et al 2002]
 - Agents for Semantic Web [Hendler 01] [Finin et al. 02] ...
 - Semantic Web browsers: Haystack [Quan,Huynh,Karger 03], Magpie [Dzbor,Domingue,Motta 03]

Need for application in real-world community context

- ▶ Provides an **enhanced semantic interface** to bug resolution messages
- ▶ Provides **message recommendations** of resources for each bug resolution message

OpenACS Bug #1338

3/05/2004 Opened by [Mark Aufflick](#)

if you don't supply a var required by ad_page_contract, you get a server error:

```
can't read "exception_count": no such variable
while executing
"lang::util::localize ${exception_count}"
...
```

[Hide recs](#)

People: lars pind
dave bauer

Code: bootstrap.tcl
gpl.html

Bug reports: Parsing error when a user doesnot insert proper HTML tags (1833)
return_url in entry-edit should be "" not "." (382)

Forums: ad_page_contract blocking less than symbol in form data (165028)
multi var validation (89072)

Realizing Dhruv for bug resolution ...

Focus on linking bug report messages to code

- A. Metadata** for community artifacts
 - *Identify potential code terms in messages*
- B. Ontologies** of the content, interactions and community of project
 - *Locate code terms in an ontology of the code*
- C. Heuristics** to automatically link related objects and artifacts
 - *Use ontology relations to identify context of code terms (cross-links)*
- D. Recommendation procedures** to suggest related artifacts during bug resolution
 - *Use multiple cross-links to generate message recommendations*

A: Identifying potential code terms

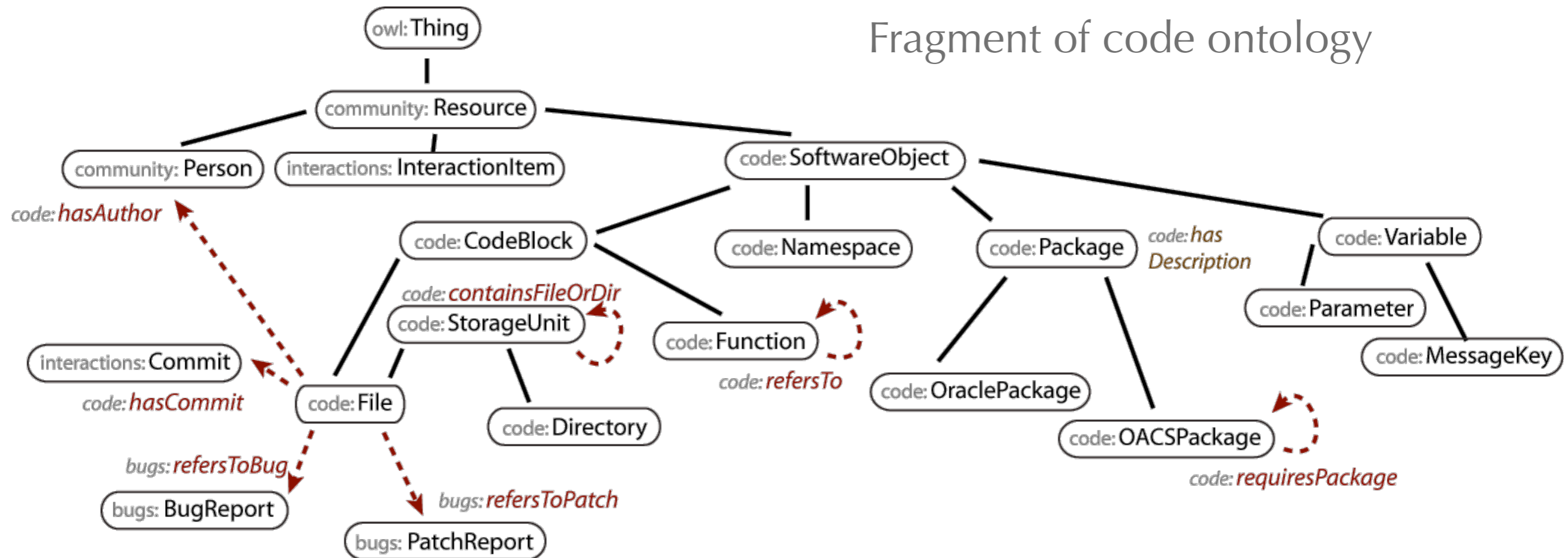
3/01/2004 Resolved (Fixed) by [Dirk Gomez](#)

The URL for a `file-storage` folder contains I18Nized strings and the tcl pages don't localize them. Furthermore some of the links were missing `@?version_id=@contents.live_revision@`. Fixed that.

References in message text

- Information extraction (IE) rules to extract code references, files, error traces
- Potential code terms indicated by
 - Programming language syntax
 - Terms with underscores, hyphens etc.
- Detecting references to code is non-trivial because messages contain code in different programming languages and often references mangled
- Focus on high recall
- Opted for generous IE, then pruned extracted terms by comparison to known software objects in code ontology

B. Locating terms in ontology



- ▶ Developed OWL DL ontologies to describe code content, interactions and community of the OpenACS project
- ▶ Relatively general: requiring modification of small percentage of concepts and relations for use in other OSS communities

B: Locating terms in code ontology

Where are the code terms found
in the software code?

Ontology with information:

- Function found in files
- Function has variables
- Function calls other function
- Function is in namespace
- Namespace is in files
- ...

Ontology facts (metadata)
captured by processing the
OpenACS code base

B: Locating terms in code ontology

Where are the code terms found in the software code?

Ontology with information:

- Function found in files
- Function has variables
- Function calls other function
- Function is in namespace
- Namespace is in files
- ...

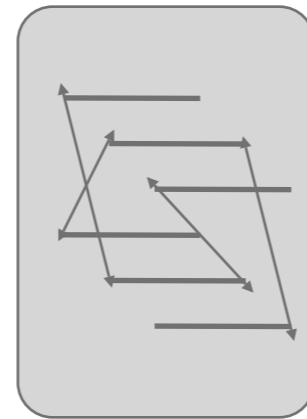
Guiding ontology:

- An Integrated Representation for Software Development and Discovery [Welty 1995]
- Semantics of software structure more relevant than semantics of code

Ontology facts (metadata) captured by processing the OpenACS code base

B: Locating terms in code ontology

Where are the code terms found in the software code?



Ontology with information:

- Function found in files
- Function has variables
- Function calls other function
- Function is in namespace
- Namespace is in files
- ...



Guiding ontology:

- An Integrated Representation for Software Development and Discovery [Welty 1995]
- Semantics of software structure more relevant than semantics of code

Ontology facts (metadata) captured by processing the OpenACS code base

C: Identifying context of code terms

- Objects and artifacts identified via code ontology
 - Found in file
 - Found as
 - Defined in file
 - Contains variables
 - Functions called
 - Has author
 - Functions in same file
 - Recent committers to the file
 - Bug reports filed for file
- Artifacts based on text similarity

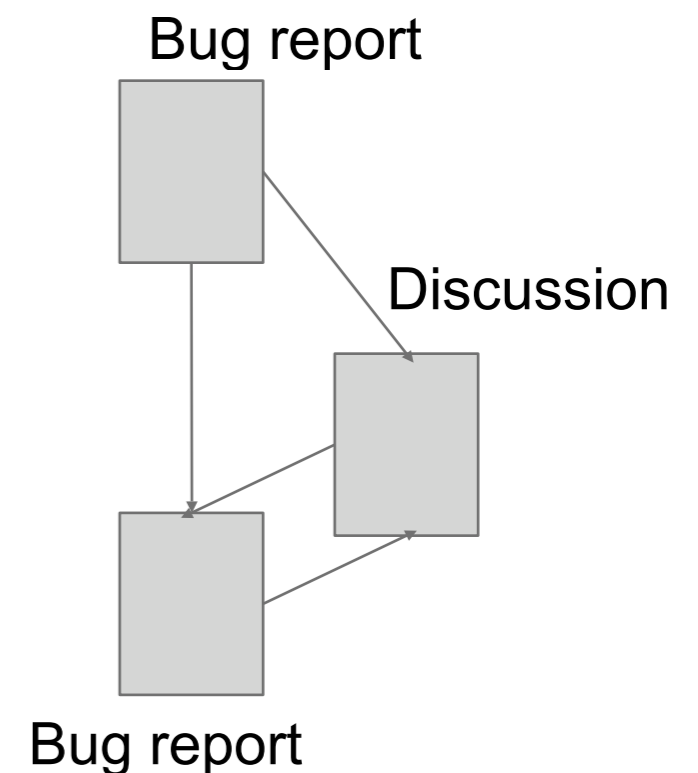
C: Identifying context from artifact references

During discussions, people naturally mention related artifacts

- Capture references as inter-artifact links
- Explicit directed links between resources

Heuristic:

- Apply to artifacts determined from ontology relations and text similarity
- Convert into undirected links and select artifacts one hop away



D. Message Recommendations

For each message

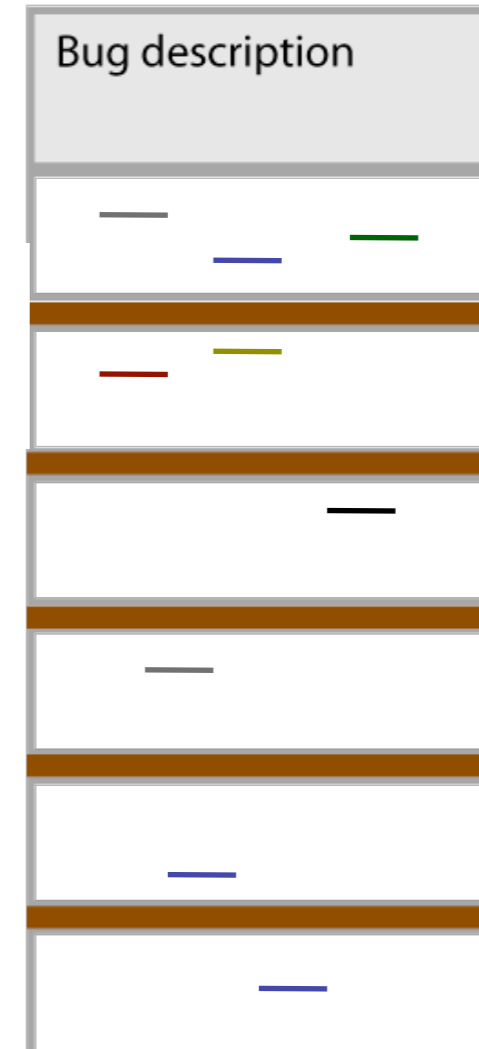
1. Use set of artifacts generated for message terms in previous step
2. Aggregate and rank artifacts, such that higher rank for artifacts generated from
 1. Multiple different terms
 2. Multiple sources

In addition:

- Higher weights for terms in bug report summary
- Include artifact references within message
- Accumulate message recommendations



Bug report



Bug Comment

Evaluation: Think-aloud

Objective: to obtain detailed feedback from OpenACS community on various aspects of Dhruv and its potential use

Participants: Veteran OpenACS community members with various leading roles in community

Study

1. OpenACS community members presented with bug report to be analysed
2. Bug report chosen, such that
 - Study participant was involved in resolution of bug report
 - Bug report resolved in distant past (> 6 months ago)
3. Participant asked to use standard bug tracker first and then use Dhruv to analyse bug report

What did the community think of Dhruv?

- Enhanced semantic interface
 - Very useful
 - Low precision of highlighted terms, but human filtering possible
 - Cross-links present useful information, especially when small in number
- Message recommendations
 - Not directly useful for bug resolution, at least in first pass
 - People recommendations good idea, but need to take area specialization into account
 - Message recommendations are relevant, but need to reduce spurious recommendations

Community suggestions for improving Dhruv

Community suggestions for improving Dhruv

```
if you don't supply a var required by ad_page_contract, you  
  
can't read "exception_count": no such variable  
  while executing  
  "lang::util::localize ${exception_count}"  
  
...
```

A. Make
machine-
generated
aspect
salient

Community suggestions for improving Dhruv

```
if you don't supply a var required by ad_page_contract, you  
  
can't read "exception_count": no such variable  
while executing  
"lang::util::localize ${exception_count}"  
  
...
```

A. Make
machine-
generated
aspect
salient

[Hide recs](#)

People: lars pind
dave bauer

Code: bootstrap.tcl
gpl.html

Bug reports: Parsing error when a user doesnot insert proper HTML tags (1833)
return_url in entry-edit should be "" not "." (382)

Forums: ad_page_contract blocking less than symbol in form data (165028)
multi var validation (89072)

B. Hide recommendations
by default

Community suggestions for improving Dhruv

```
if you don't supply a var required by ad_page_contract, you  
  
can't read "exception_count": no such variable  
while executing  
"lang::util::localize ${exception_count}"  
  
...
```

A. Make
machine-
generated
aspect
salient

[Hide recs](#)

People: lars pind
dave bauer

Code: bootstrap.tcl
gpl.html

Bug reports: Parsing error when a user doesnot insert proper HTML tags (1833)
return_url in entry-edit should be "" not "." (382)

Forums: ad_page_contract blocking less than symbol in form data (165028)
multi var validation (89072)

B. Hide recommendations
by default

ad_page_contract

```
owl:Thing  
community:Resource  
code:SoftwareObject  
code:CodeBlock  
code:Function
```

Metadata

Defined in file:

[acs-tcl/tcl/tcl-documentation-procs.tcl](#)

Defined in namespace:

Authored by:

Bryan Quinn
Jon Salz
Lars Pind
Yonatan Feldman

Other functions in same file:

```
ad_complain  
ad_complaints_count  
ad_complaints_get_list  
ad_complaints_init  
ad_complaints_parse_error_strings  
ad_complaints_with_key  
ad_conn  
ad_footer  
ad_header_with_extra_stuff  
ad_html_security_check  
ad_page_contract  
ad_page_contract_complaints
```

Community suggestions for improving Dhruv

```
if you don't supply a var required by ad_page_contract, you  
  
can't read "exception_count": no such variable  
while executing  
"lang::util::localize ${exception_count}"  
  
...
```

A. Make
machine-
generated
aspect
salient

B. Hide recommendations
by default

[Hide recs](#)

People: lars pind
dave bauer

Code: bootstrap.tcl
gpl.html

Bug reports: Parsing error when a user doesnot insert proper HTML tags (1833)
return_url in entry-edit should be "" not "." (382)

Forums: ad_page_contract blocking less than symbol in form data (165028)
multi var validation (89072)

ad_page_contract

~~owl:Thing
community:Resource
code:SoftwareObject
code:CodeBlock
code:Function~~

Metadata

Defined in file:

[acs-tcl/tcl/tcl-documentation-procs.tcl](#)

Defined in namespace:

Authored by:

Bryan Quinn
Jon Salz
Lars Pind
Yonatan Feldman

Other functions in same file:

ad_complain
ad_complaints_count
ad_complaints_get_list
ad_complaints_init
ad_complaints_parse_error_strings
ad_complaints_with_key
ad_conn
ad_footer
ad_header_with_extra_stuff
ad_html_security_check
ad_page_contract
ad_page_contract_complaints

C. Remove
ontology
information

Community suggestions for improving Dhruv

```
if you don't supply a var required by ad_page_contract, you  
  
can't read "exception_count": no such variable  
while executing  
"lang::util::localize ${exception_count}"  
  
...
```

A. Make
machine-
generated
aspect
salient

ad_page_contract

~~owl:Thing
community:Resource
code:SoftwareObject
code:CodeBlock
code:Function~~

Metadata

Defined in file:

[acs-tcl/tcl/tcl-documentation-procs.tcl](#)

Defined in namespace:

Authored by:

Bryan Quinn
Jon Salz
Lars Pind
Yonatan Feldman

Other functions in same file:

ad_complain
ad_complaints_count
ad_complaints_get_list
ad_complaints_init
ad_complaints_parse_error_strings
ad_complaints_with_key
ad_conn
ad_footer
ad_header_with_extra_stuff
ad_html_security_check
ad_page_contract
ad_page_contract_complaints

C. Remove
ontology
information

[Hide recs](#)

People: lars pind
dave bauer

Code: bootstrap.tcl
gpl.html

Bug reports: Parsing error when a user doesnot insert proper HTML tags (1833)
return_url in entry-edit should be "" not "." (382)

Forums: ad_page_contract blocking less than symbol in form data (165028)
multi var validation (89072)

B. Hide recommendations
by default

D. Reduce
length

Community suggestions for improving Dhruv

```
if you don't supply a var required by ad_page_contract, you  
  
can't read "exception_count": no such variable  
while executing  
"lang::util::localize ${exception_count}"  
  
...
```

A. Make
machine-
generated
aspect
salient

ad_page_contract

~~owl:Thing
community:Resource
code:SoftwareObject
code:CodeBlock
code:Function~~

Metadata

Defined in file:

[acs-tcl/tcl/tcl-documentation-procs.tcl](#)

Defined in namespace:

Authored by:

Bryan Quinn
Jon Salz
Lars Pind
Yonatan Feldman

Other functions in same file:

ad_complain
ad_complaints_count
ad_complaints_get_list
ad_complaints_init
ad_complaints_parse_error_strings
ad_complaints_with_key
ad_conn
ad_footer
ad_header_with_extra_stuff
ad_html_security_check
ad_page_contract
ad_page_contract_complaints

C. Remove
ontology
information

B. Hide recommendations
by default

[Hide recs](#)

People: lars pind
dave bauer

Code: bootstrap.tcl
gpl.html

Bug reports: Parsing error when a user doesnot insert proper HTML tags (1833)
return_url in entry-edit should be "" not "." (382)

Forums: ad_page_contract blocking less than symbol in form data (165028)
multi var validation (89072)

E. Direct interaction with Dhruv

D. Reduce
length

Community suggestions for improving Dhruv

```
if you don't supply a var required by ad_page_contract, you  
  
can't read "exception_count": no such variable  
while executing  
"lang::util::localize ${exception_count}"  
  
...
```

A. Make
machine-
generated
aspect
salient

ad_page_contract

~~owl:Thing
community:Resource
code:SoftwareObject
code:CodeBlock
code:Function~~

Metadata

Defined in file:

[acs-tcl/tcl/tcl-documentation-procs.tcl](#)

Defined in namespace:

Authored by:

Bryan Quinn
Jon Salz
Lars Pind
Yonatan Feldman

Other functions in same file:

ad_complain
ad_complaints_count
ad_complaints_get_list
ad_complaints_init
ad_complaints_parse_error_strings
ad_complaints_with_key
ad_conn
ad_footer
ad_header_with_extra_stuff
ad_html_security_check
ad_page_contract
ad_page_contract_complaints

C. Remove
ontology
information

B. Hide recommendations
by default

[Hide recs](#)

People: lars pind
dave bauer

Code: bootstrap.tcl
gpl.html

Bug reports: Parsing error when a user doesnot insert proper HTML tags (1833)
return_url in entry-edit should be "" not "." (382)

Forums: ad_page_contract blocking less than symbol in form data (165028)
multi var validation (89072)

E. Direct interaction with Dhruv

F. Elicit more information from users
at bug reporting time

D. Reduce
length

Dhruv fits into natural way of working in OpenACS

“I would say that it fits into that natural way of trying to grab at things outwards from a few little kernels ... trying to take this one and where can I go with this particular piece of code or this comment. It would jive with my kind of approach.”

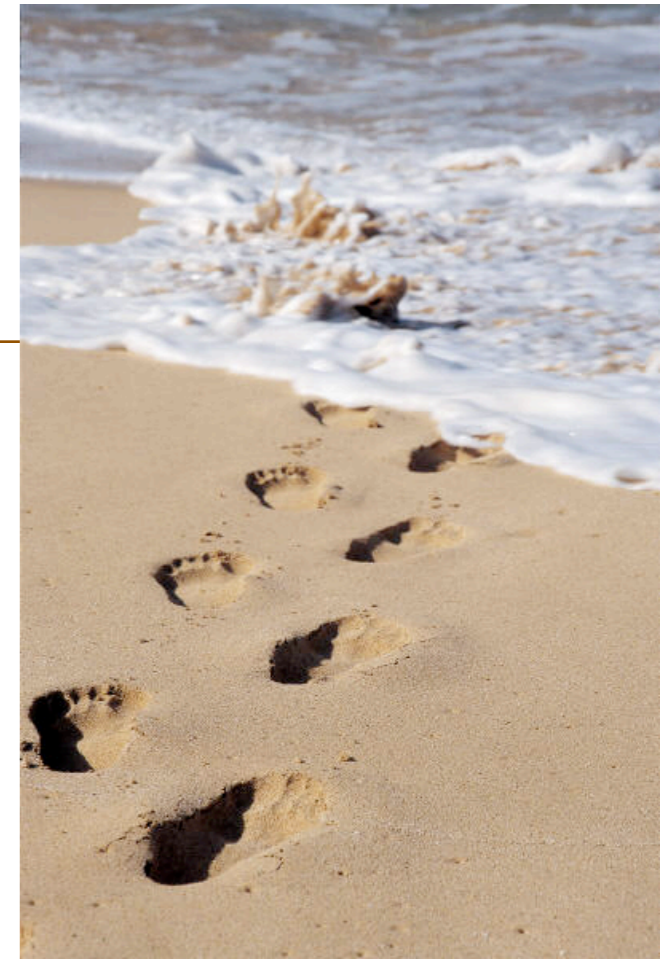
“If someone is ... trying to learn a little bit more about a particular bug, you’ve provided many different pathways to code that’s related so you’ve learnt a little bit about different related procedures just by clicking around on those links.”

Findings/Issues

- ▶ Large amount of complex data generated by community ...
expressivity/performance tradeoff
- ▶ Reasoning underutilised in Dhruv, partly due to size and
complexity of data
- ▶ Community ontologies need real-world validation and usage
- ▶ Dhruv models OSS community as closed information space
 - appropriate for current usage scenarios
 - OWL more useful for linking such isolated 'Semantic Web
communities'?
- ▶ Involve community in creation of Semantic Web
 - OSS community is technically sophisticated and highly
invested in improving their own communities and
development processes
 - Building Dhruv from the 'outside' makes it light-weight and
requires little change to existing systems for deployment

Next steps ...

- ▶ Deployment in real-world contexts
 - Field study of system in OpenACS community to understand how people act on suggestions
 - Improve individual components of Dhruv
- ▶ Application to other contexts
 - Other OSS activities such as code comprehension
 - Other OSS communities
 - Online professional communities in education and scientific research
- ▶ Framework for transitioning Web communities to Semantic Web communities



where does
Web 2.0
come in
here?

Thank you!

Questions?

Anupriya Ankolekar
anupriya@aifb.uni-karlsruhe.de