Network Working Group
Request for Comments: 2972
Category: Informational

N. Popp
RealNames Corporation
M. Mealling
Network Solutions
L. Masinter
AT&T Labs
K. Sollins
MIT

October 2000

Context and Goals for Common Name Resolution

Status of this Memo

Copyright Notice

Abstract

This document establishes the context and goals for a Common Name Resolution Protocol. It defines the terminology used concerning a "Common Name" and how one might be "resolved", and establishes the distinction between "resolution" and more elaborate search mechanisms. It establishes some expected contexts for use of Common Name Resolution, and the criteria for evaluating a successful protocol. It also analyzes the various motivations that would cause services to provide Common Name resolution for both public, private and commercial use.

This document is intended as input to the formation of a Common Name Resolution Protocol working group. Please send any comments to cnrp-ietf@lists.internic.net. To review the mail archives, see <http://lists.internic.net/archives/cnrp-ietf.html>

1. Introduction

People often refer to things in the real world by a common name or phrase, e.g., a trade name, company name, or a book title. These names are sometimes easier for people to remember and enter than URLs; many people consider URLs hard to remember or type. Furthermore, because of the limited syntax of URLs, companies and individuals are finding that the ones that might be most reasonable

for their resources are already being used elsewhere and therefore unavailable. Common names are not URIs (Uniform Resource Identifiers) in that they lack the syntactic structure imposed by URIs; furthermore, unlike URNs, there is no requirement of uniqueness or persistence of the association between a common name and a resource. These common names are expected to be used primarily by humans (as opposed to machine agents).

Common name "resolution" is a process of mapping from common names to Internet resources; a Common Name Resolution Protocol (CNRP) is a network protocol used in such a process.

A useful analogy for understanding the purpose and scope of common names, and CNRP, are everyday (human language) dictionaries. These cover a given language (namespace) -- perhaps a spoken language, or some specific subset (e.g., technical terms, etc). Some dictionaries give definitions, others give translations (e.g., to other languages). Different entities publish dictionaries that cover the same language -- e.g., Larousse and Collins can both publish French-language dictionaries. Thus, the dictionary publisher is the analog to the resolution service provider -- the service can provide a value-add and build up name recognition for itself, but does not impede other entities from providing definitions for precisely the same strings in the language.

Services are arising that offer a mapping from common names to Internet resources (e.g., as identified by a URI). These services often resolve common name categories such as company names, trade names, or common keywords. Thus, such a resolution service may operate in one or a small number of categories or domains, or may expect the client to limit the resolution scope to a limited number of categories or domains. For example, the phrase "Internet Engineering Task Force" is a common name in the "organization" category, as is "Moby Dick" in the book category. A single common name may be associated with different data records, and more than one resolution service is expected to exist. Any common name may be used in any resolution service.

Two classes of clients of such services are being built: browser improvements and web accessible front-end services. Browser enhancements modify the "open" or "address" field of a browser so that a common name can be entered instead of a URL. Internet search sites integrate common name resolution services as a complement to search. In both cases, these may be clients of back-end resolution services. In the browser case, the browser must talk to a service that will resolve the common name. The search sites are accessed via

a browser. In some cases, the search site may also be the back-end resolution service, but in others, the search site is a front-end to a collection of back-end services.

This effort is about the creation of a protocol for client applications to communicate with common name resolution services, as exemplified in both the browser enhancement and search site paradigms. Although the protocol's primary function is resolution, it is intended to address the issues of internationalization, authentication and privacy as well. Name resolution services are not generic search services and thus do not need to provide complex Boolean query, relevance ranking or similar capabilities. The protocol is expected to be a simple, minimal interoperable core. Mechanisms for extension will be provided, so that additional capabilities can be added later.

Several other issues, while of importance to the deployment of common name resolution services, are outside of the resolution protocol itself and are not in the initial scope of the proposed effort. These include discovery and selection of resolution service providers, administration of resolution services, name registration, name ownership, and methods for creating, identifying or insuring unique common names.

2. Key Goals for a Common Name Resolution Protocol

The key deliverable is a protocol for parameterized resolution. "Resolution" is defined as the retrieval of data associated (a priori) with descriptors that match the input request. "Parameterized" means the ability to have a multi-component descriptor both as part of the query and the response. These descriptors are attribute-value pairs. They are not required to provide unique identification, therefore 0 or more records may be returned to meet a specific input query. The protocol will define:

- client requests/server responses to identify the specific parameters accepted and/or required on input requests
- client request/server responses to identify properties to be returned in the result set
- expression of parameterized input query
- expression of result sets
- standard expression of error conditions

To avoid creating a general search protocol with unbounded complexity, and to keep the protocol simple enough so that different implementations will have similar behavior, the resolution protocol should be limited to sub-string matches against parameter values. To support full internationalization, UTF-8 encoding of strings and sub-strings is preferred.

In addition, the working group should define one sample service based on this protocol -- the resolution of so-called "common names", or resolution of non-unique, registered strings to resource descriptions.

3. CNRP goals

The goal of CNRP is to create a lightweight search protocol with a simple query interface, with a focus on making the common case of substring search with a single result most efficient. In addition, efficient support for keyed value search is important. Each key is a named meta property of the resource (e.g. category, language, geographical region.). Some of these properties could be standardized (e.g. the common name property). The goal is to support partial specification of query parameters and even partial and fuzzy matches on names. CNRP is intended to be simpler than LDAP for simple applications.

Besides simplicity, the CNRP protocol should be consistent with efficient implementation of a simple and intuitive user interface. The emphasis on the common name as the common denominator to find a wide range of resources reduces the UI to its minimal expression (the user types a few words in a text box and presses enter).

CNRP should provide interoperability with multiple common name databases (section 4 presents many examples of such databases). The query interface should be extensible and customizable to the specific needs of a specific type of resolution service. However, the need for interoperability across databases and resolution services combined with the need to ensure the scalability of search (across millions of names from multiple providers) have lead this group to consider the explicit requirement of supporting categories in CNRP. This requirement is discussed further in section 5.

4. Example of common name namespaces

Commercial companies have already developed and deployed common name resolution services such as RealNames (http://www.realnames.com) and NetWords (http://www.netword.com). These commercial implementations are mainly focused on trade names, such as company names, brands and

trademarks. These services constitute a concrete example of common name namespaces implementation and are useful to understand the scope of the CNRP effort.

CNRP is also directly targeted at directory service providers. CNRP is relevant to these services to increase their reach through integration into larger Web sites such as the search portals. For example, IAtlas has developed a directory service for businesses that it distributes through its Web site and Inktomi. IAtlas could immediately leverage CNRP to distribute their service through their external distribution partners.

Directory services must not be confused with search engines. Directory services use highly structured information to identify a resource. This information is external to the actual resource and is called metadata. In contrast, search engines mainly rely on the content of the resource (e.g. the text of a Web page).

CNRP plays well with directory services that present a critical piece of information about the resource in the form of a textual identifier, a title or a terse description (the common name). Numerous examples come instantly to mind: company names, book titles, people names, songs, ISBNs, and social security numbers. In all cases, the common name is the natural property for users to lookup the resource. The common name is always simple and intuitive: it has no syntax, it is multilingual, memorable and can often be guessed.

The following list is intended to put in prospective the wide range of applications for CNRP:

- Business directories (SEC, NASDAQ, E*Trade, .). The resource is company information (address, products, SEC filings, stock quotes, etc.). The common name is the company name.

- White pages (BigFoot, WhoWhere, Switchboard, ...): The resource a person (current address, telephone numbers, email addresses, employer, ...). The common name is a last name, a telephone number or an email address.

- E-commerce directories: The resource is a product for sale (car, house, furniture, actually almost any type of consumption item). The common name is a brand name or a description.

- Publishing directories: The resource is one of many things: a book, a poem, a CD, an MP3 download. The common name is an ISBN, a song title, an artist’s name. The common name is typically the title of a publication.

- Entertainment directories: The resource is an event (a movie, a concert, a TV show). The common name is the name or a description for the event, the movie title, a rock band name, a show.

- Yellow pages services: Here again, the resource can be very diverse: a house for sale, a restaurant, a car dealership or other type of establishment or service that can be found in the traditional yellow pages. The common name can be a street address, the name of a business, or a description.

- News feeds: The resource is a press article. The common name is the headline.

- Vertical directories: the DNS TLD categories, the ISO country codes.

5. Private and public namespaces

A set of common names within a category (books, news, businesses, etc.) is called a common name "namespace". The term "namespace" only refers to the set of names. It does not encompass the bindings or associations between a name and data about the name (such as a resource, identified by a URI). Such bindings might be created and maintained by a common name resolution services. Resolution services may create binding that are relevant for the type of service that they offer.

It is useful to distinguish between "private" and "public" namespaces. A namespace is private if owned by an authority that controls the right to assign the names. A namespace is private even if the right to assign those names is held by a neutral party.

A namespace is public when not controlled by any single authority or resolution provider. Assignment of the names is distributed. However, it is reasonable to expect that people who assign names will tend to pick names that have a minimum of collisions. For some of these namespaces, there will even be mechanisms to discourage duplicate assignment, but all of them are inherently ambiguous. Public namespaces are not controlled. Examples of public namespaces are:

- Titles of books, movies, songs, poems, short stories, plays, or compilations
- Place names
- Street names
- People’s names

Because these namespaces are unbounded and open to any types of name assignment, they will have scalability problems. To support these namespaces, CNRP must provide at least one standard mechanism to filter a large list of related results. A filtering mechanism must allow the user to narrow the search further down to a smaller result set, because the common name alone may not be enough.

One possible search filter is related to the notion of categories. Because categories create a structure to organize named resources, large resolution services are likely to support some sort of categorization system (whether flat or hierarchical). Although categories constitute an efficient search filter, defining standard vocabularies for common name categories is beyond the scope of the protocol design. The protocol design for CNRP should not require a standardized taxonomy for categories in order to be effective. For example, CNRP resolution could use free-form keywords; the end-user would use these keywords as part of the query. Each service would then be responsible for mapping the keywords to zero, one or many categories in their own classification. The keywords would remain classification independent and different services could use different categorization schemes without compromising interoperability. It would then be up to the service to provide its own mapping. For example, let us assume that one namespace is resolving names under the category: "Hobby & Interests > collecting > antique > books". Assume that a second namespace has decided to organize the names of similar resources under the classification: "Arts > Humanities > Literature > History of Books and Printing > antiques". Although the two taxonomies are different, a CNRP query specifying category_keywords = "antique books" would allow each service to identify the appropriate category. This mechanism may ensure that the two result lists are small and coherent enough to be merged into one unique result set. It is important to note that this approach would work whether the classification is hierarchical or not.

Although this suggestion has merit, it is fair to say that it remains unproven. In particular, it is unclear that the category_keywords property would guarantee full interoperability across resolution services. In any case, free form keywords for specifying categories is just one of several possible ways of limiting the scope of a query. Although the specific mechanisms are not agreed upon a this time, CNRP will provide at least one standard mechanism for limiting scope.

6. Distributors/integrators of common name resolution services

We anticipate two main categories of distributors for common namespaces. The first category is made of the Web portals such as search engines (Yahoo, MSN, Lycos, Infoseek, AltaVista, ...). A

common name resolution service will typically address only one very specialized aspect of search (company names or book titles or people names, ..). This type of focused lookup service is a useful complement to generic search. Hence, portals are likely to integrate several types of common name services. CNRP solves the difficult problem of integrating multiple external independent services within one Web site. Today, the lack of standardization in performance requirements and query interface leads to loose integration (co-branded pages hosted on virtual domains) or maintenance problems (periodical data dumps). CNRP is aimed at solving some of these issues. CNRP facilitates the deployment of embedded services by creating a common interface to all common name services.

The second category of distributors is made of the Web browser companies. Netscape's smart browsing (http://home.netscape.com/communicator/v4.5/index.html#smart) and Microsoft's IE5 auto-search features (http://www.microsoft.com/windows/Ie/Features/AutoSearch/default.asp) demonstrate that the two dominant Web browser companies understand the value of navigation and search from the command line of the browser. It is very clear how this command line could be used as the main user interface to common name resolution services through CNRP. In many ways, it is actually the most natural user interface to resolve a common name. For this strategic component of the browser's user interface to remain truly open to all common name resolution services, it is key that there exists a standard resolution protocol (and a service discovery mechanism). CNRP will give users access to the largest selection of services and providers and the ability to select a specific resolution service over another. To preserve the user from proprietary implementations, the existence of CNRP is a prerequisite.

7. Example of cost recovery models for maintenance of namespaces

The following discussion of possible business models for common name namespaces is intended to prove that they are commercially viable, hence that CNRP will be used in the market place. This section presents 5 different cost recovery models.

a. Licensing the lookup service

In such model, the owner of the database owner licenses the data and the resolution service to a portal. This is a proven model. For example, Looksmart (a directory service) recently licensed all their data to MSN. Another possibility is to sell access to the service directly to the user. For some vertical type of common

names service (e.g. patent search), it is also conceivable that a specific type of users (e.g., lawyers) would be willing to pay for accessing a precise resolution service.

b. Sharing revenue generated by banner advertising

In this model, the database owner licenses his infrastructure (data and resolution service) to a portal. Prepaid banner ads are placed on the result pages. The revenue is shared between the resolution service provider and the portal that hosts the pages.

c. Selling the names (charge the customer a fee for subscribing a name)

This is a proven business model as well (NSI, GOTO, RealNames, Netword, for of the name has a large user reach (search engines sell keywords for instance).

d. Value added service

Another model is to build a common name as a free added value service in order to make a core service more compelling to users. For example, Amazon.com could create a common name namespace of book titles and make it freely available to its users. Amazon.com would not make any money from the resolution service per se. However, it would indirectly since the service would help the users find hence buy more books from Amazon.com.

e. "Some-strings-attached" free names

A namespace may give users a name for free in exchange for something else (capturing the user’s profile that can be sold to merchants, capturing the user’s email address in order to send advertising emails, etc.).

8. Security and Intellectual Property Rights Considerations

This document describes the goals of a system for multi-valued Internet identifiers. This document does not discuss resolution; thus questions of secure or authenticated resolution mechanisms are out of scope. It does not address means of validating the integrity or authenticating the source or provenance of Common Names. Issues regarding intellectual property rights associated with objects identified by the various Common Names are also beyond the scope of this document, as are questions about rights to the databases that might be used to construct resolvers.

9. Authors' Addresses

Larry Masinter
AT&T Labs
75 Willow Road
Menlo Park, CA 94025

Phone: +1 650 463 7059
EMail: LMM@acm.org
http://larry.masinter.net

Michael Mealling
Network Solutions
505 Huntmar Park Drive
Herndon, VA 22070

Phone: (770) 935-5492
Fax: (703) 742-9552
EMail: michaelm@netsol.com

Nicolas Popp
RealNames Corporation
2 Circle Star Way
San Carlos, CA 94070-1350

Phone: 1-650-298-5549
EMail: nico@realnames.com

Karen Sollins
MIT Laboratory for Computer Science
545 Technology Sq.
Cambridge, MA 02139

Phone: +1 617 253 6006
EMail: sollins@lcs.mit.edu

10. Full Copyright Statement

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.