Search-based Contextual Routing via Isomorphic Uniform Resource Identifiers

Introduction

It is the intent of this paper to describe a system whereby a strong isomorphic relationship[1] is retained between a data resource request and the data returned by a resource via a search-based method to derive contextual meaning from compliant Uniform Resource Identifiers (URI)[2]. URIs retaining a strong isomorphic relationship between a resource request and returned data as described within the scope of this paper will be referred to as Isomorphic Uniform Resource Identifiers (IURIs). Further, a primary method used to retain a strong isomorphic relationship between IURIs and the returned data will be referred to as Search-based Contextual Routing (SbCR).

Resource Request and Return Relationships

A resource request (REQ) as a URI and a resource return (RET) as data will express an isomorphic exchange where elements of a URI can be correspondingly mapped unto the returned data structure and where a data's structure can be correspondingly mapped onto the elements of a URI – simply, the process must be reversible. Many transformations between the original REQ and RET may occur, though a corresponding mapping must largely be retained to qualify as an isomorphic exchange.

Today, many such corresponding mappings of URIs and data are implemented on the internet -- however, the strength of isomorphism and the methods in which the data is mapped varies wildly. The varying degrees of isomorphism and data mapping, though usually URI-compliant, make it difficult to reliably request data from a given resource without intimate knowledge of how that resources is structured.

An example of this would be a typical web search done in both Google and Yahoo's search engines:

Search term = *Isomorphic*

---

[1] "*The word "isomorphism" applies when two complex structures can be mapped onto each other, in such a way that to each part of one structure there is a corresponding part in the other structure, where "corresponding" means that the two parts play similar roles in their respective structures.*" (Gödel, Escher, Bach, p. 49 / Douglas Hofstadter)

[2] *A URI can be further classified as a locator, a name, or both. The term "Uniform Resource Locator" (URL) refers to the subset of URIs that, in addition to identifying a resource, provide a means of locating the resource by describing its primary access mechanism (e.g., its network "location"). The term "Uniform Resource Name" (URN) has been used historically to refer to both URIs under the "urn" scheme [RFC2141], which are required to remain globally unique and persistent even when the resource ceases to exist or becomes unavailable, and to any other URI with the properties of a name.*

An individual scheme does not have to be classified as being just one of "name" or "locator". Instances of URIs from any given scheme may have the characteristics of names or locators or both, often depending on the persistence and care in the assignment of identifiers by the naming authority, rather than on any quality of the scheme. Future specifications and related documentation should use the general term "URI" rather than the more restrictive terms "URL" and "URN" [RFC3305]. (1.1.3 - RFC 3986 / Berners-Lee, et al 2005)

URI expression (Google): http://www.google.com/search?q=isomorphic[3]
URI expression (Yahoo!): http://search.yahoo.com/search?&p=isomorphic[4]

Though both have very similar URI structures and return resources relevant to the search term *Isomorphic,* the URI structures are different enough to make it difficult to know the search URI structure of both sites by simply searching one. Specifically, Google chooses the variable 'q' which presumably abbreviates the word: query, while Yahoo! uses the variable 'p' which probably stands for: phrase. Both URIs do quite a good job at declaring what it is one would expect to find upon following the links, mainly a 'search' for query/phrase 'isomorphic' however it is common on the internet for this clear declaration to break down such as when browsing a product online. A URI of the following nature, taken from MSN Shopping is not uncommon amongst e-commerce websites:

http://shopping.msn.com/results/shp/?bCatId=7042[5]

It would be impossible to know that that URI represents a listing of men's watches without either visiting the page and parsing its contents or knowing its internal mapping structure, as such where it would be known that (CatId=7042) = "Watches for Men." To contrast, a URI for Yahoo! Shopping's men's watches:

http://shopping.yahoo.com/s:Dress%20Watches:4157-Gender=Men%27s:browsename=Men%27s%20Dress%20Watches:refspaceid=96181141;_ylt=An1dYuytOnpJRSPr7khHqw86FawC;_ylu=X3oDMTA4MmQwaHRuBHNlYwN2bmF2[6]

From this URI we can derive much more meaning, specifically 'Dress Watches' for 'Gender=Men,' however this URI proves to be irreproducible by any other system or persons to access the same data without knowing in advance the other portions of the URI specific to Yahoo! Shopping as a resource.

Nonetheless, it can be said that the Yahoo! Shopping URI example retains a *stronger* isomorphism than the MSN Shopping URI example.

Weak and Strong Isomorphic URIs

Weak: All REQ/RET exchanges, by their very nature, contain weak forms of isomorphism. The more information that must be supplied to make the mapping from a given URI to some data; the weaker its isomorphism is expressed.

Given the examples thus far, it appears the MSN Shopping example contains the *weakest* (most knowledge-dependent) isomorphism:

http://shopping.msn.com/results/shp/?bCatId=7042

From this URI we can derive:

[3] http://www.google.com (2006)
[4] http://search.yahoo.com (2006)
[5] http://shopping.msn.com (2006)
[6] http://shopping.yahoo.com (2006)

- It comes from the 'com'mercial TLD-space
    - From a domain name called 'msn'
        - With a sub-domain called 'shopping'
            - Containing a folder called 'results'
                - That contains a sub-folder called 'shp'
            - And a variable called 'bCatId'
                - With a value of '7042'

However, from that MSN Shopping URI we cannot derive that it contains watches for men, nor are we sure that 7042 +/- 1 will yield data similar to men's watches (in fact 7041 & 7043 returns no results). We can strengthen the isomorphic relationship between this URI and its resulting data only by having advance knowledge specific to MSN Shopping, such as a table listing all of the possible values of bCatId and their corresponding labels.

Is it possible to strengthen the isomorphic relationship of MSN Shopping URIs and the resulting data without advanced knowledge of MSN Shopping's structure while still preserving the usability or reproducibility of its URIs? In other words, is it possible to express a stronger isomorphism between MSN Shopping URIs and product listings?

Strong: A strong isomorphic exchange would be characterized as one that returns data which closely matches the originating request while requiring a minimal amount of inside information; the easier it is for a hypothetical viewer to infer the RET for a given REQ, the stronger the apparent isomorphism.

Examples of IURIs:

a) Strong: "shopping.msn.com/watches/men"
b) Stronger: "shopping.msn.com/watches-for-men"
c) Strongest/friendliest of the three (not URI): "msn shopping watches for men"

Notice (a), it is quite apparent that the 'shopping' section of 'msn' will return the 'men' section of 'watches'. Example (b) provides the same level of information in a more English-natural way by using 'for' in place of the forward slash '/'.

While the focus is URI-compliant REQs, it is interesting to note the resemblance (c) has to a typical search string. This is notable because the resemblance is not mere coincidence. Indeed, navigating the internet via search engines like Google has become commonplace. However, navigating the internet via search engines, while useful for generalized sets of RETs as required by typical search behavior, would not be viable as a mechanism for transforming REQs into RETs at the site level.

Even Google's quest to index the World's information is limited by the bandwidth of its capture, index and display capability while an individual websites' exposure is limited to its relevancy and popularity. Indeed in most cases, only the source contains enough information to properly convert REQs into sufficient RETs from the source.

A search engine will most likely construct a sufficient RET of links to websites for a given REQ, however the generalized search engine can do no more at that point, it is the duty of the individual websites to receive the handoff and convert the REQ into a proper RET.

How might a website construct a RET which will retain an isomorphic relationship with the originating REQ, be it directly or via a search engine hand-off?

Search-based Contextual Routing (SbCR)

SbCR is a standardized approach to mapping a REQ to a given data-structure via a sequence of correlated searches to derive unique RETs which preserve the isomorphism with the REQs, this process may in turn be referred to as a method of Isomorphic Information Retrieval (IIR).

SbCR is a method of Isomorphic Information Retrieval which relies on a few assumptions:

1. There is a more apparent mapping, or stronger isomorphism, between REQs and RETs that share data over those that do not share data
2. The more data a given REQ contains, the more unique the RET will be and the less potential data the RET will contain
3. The less data a given REQ contains, the less unique the RET will be and the more potential data a RET will contain

SbCR, in principle, is really quite simple and the benefits become clearer when viewed in terms of its isomorphic qualities.

The Isomorphic Relationship Between Human Desire and Information Retrieval

The primary communication mechanism for humans is language. Every language is compiled of words and we organize these words into various collections, such as dictionaries, to establish and maintain the consistency of our languages.

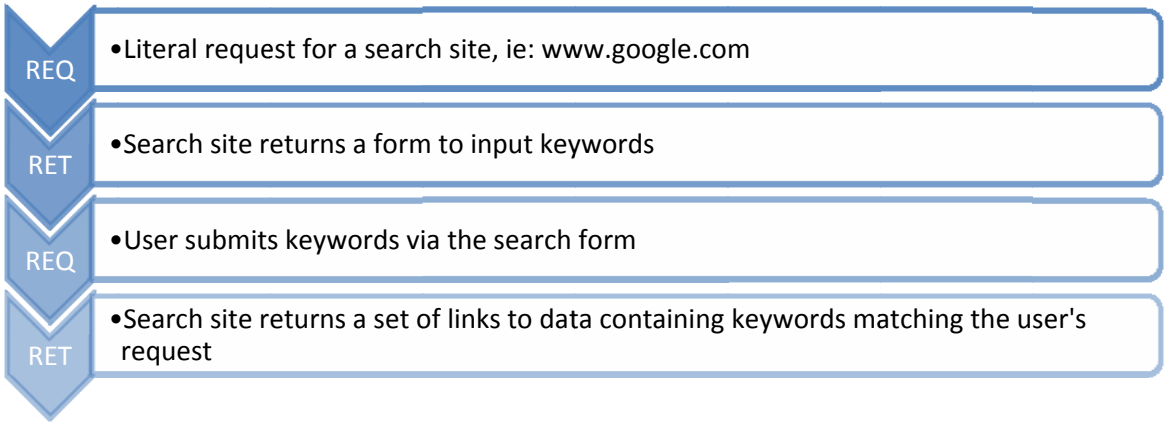
Since language is the primary technology for human communication it seems only natural to construct words and phrases which are used to identify or "key" a desired *something*, indeed this may be one of the chief reasons for the creation of language:

Communiqué	**Keywords**
"Please hand me that blue ball…"	(blue ball)
"I am looking for a short man with blonde hair…"	(short man blonde hair)
"Give me the numbers for last week's sales…"	(numbers last week's sales)
"I would like to buy a new watch…"	(buy new watch)

This derived set of keywords extends in proportion to the length and uniqueness of a given set of data. Where a few keywords may be derived from one of the above common phrases, many more keywords will be derived from a paragraph, such as taken from Verne's "From the Earth to the Moon."[7]

[7] Jules Verne, ISBN: 1598185551, http://isbndb.com/d/book/from_the_earth_to_the_moon_a03.html

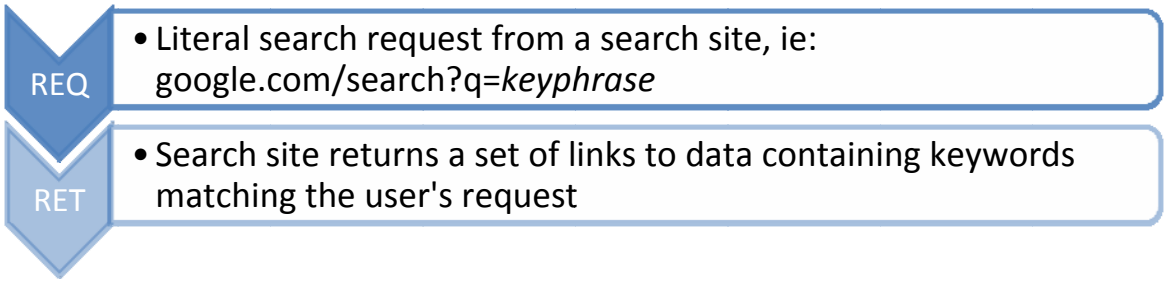
Hence in a pure data environment, like the internet, it is natural to use keywords and phrases to locate sets of data. This mode of Information Retrieval is most prevalent throughout the Internet; typically a user literally requests a search site (ie: www.google.com), the search site returns a form which allows the user to input keywords that will most likely describe the desired results, the user submits the keywords, the search site replies with a set of links to data that can derive the keywords that the user has entered.



It should be briefly noted that this picture is made a bit more complex by the notion that the links returned by the search site, may not point to data that can derive the keywords entered by the user but rather, related keywords.

This additional complexity can be referred to as Context Mediation (CM) and can be deployed in such cases where the user has entered a common misspelling, CM would jump in and either return a link to a search with the proper spelling of the keyword or assume the proper spelling and return links to data as normal. However, CM will largely be assumed to be irrelevant to the conclusions of this paper.

Furthermore in the scope of this paper, the process as stated in the above graph is too verbose and should thus be reduced:



The keyphrase variable has been italicized to illustrate the importance of that specific portion of the REQ. In this case, the portions of the string containing the TLD and domain are important in so far as matching the user's desire to [in this case] conduct a search using Google specifically.

However, 'searching' at 'Google' is merely the requested user-interface for accessing data associated with the keyphrase in the REQ. Therefore, the keyphrase is more active in determining the strength of

the isomorphism between the REQ and the RET. This is not to say that the requested user-interface is any less important, merely more ambiguous in connection with the user's desire.

The keyphrase is the active ingredient in the isomorphism between the REQ to Google and the RET of a set of links relevant to the keyphrase. While at the same time, the keyphrase being a pseudo language phrase, is most active in the isomorphism between the REQ to Google and the user's desire that ultimately triggered the creation of the REQ.

Which in the case of a simple web search, this idea seems ordinary and quite plain to see. As an example, imagine a search engine results page which is currently displaying a list of links. If most of the link's summaries described an animatronic-duck then it would be safe to assume the search was probably about animatronics and perhaps more specifically of the duck variant. Of course the conclusion could be verified by simply observing the keyphrase/search-term used.

At this point, by capturing nothing more than the RET of the REQ that produced this page, an inference can be drawn from the RET to the human desire that constructed the REQ specifically, via the isomorphic strength of the keyphrase.

Again, this insight appears quite dull and exceedingly obvious, however this stands less true for the larger mass of user-interfaces on the internet. In the last example, a search interface was chosen to handle the desired REQ, yet there are many other types of user-interfaces on the internet and each vary drastically in the isomorphic relationship between a REQ of the user-interface and the RET constructed by the interface.

The next section will focus on how it might be possible to cascade REQs into various user-interfaces with increasing fidelity and granularity by maintaining the strongest possible isomorphism at each exchange with the originating REQ.