慶應義塾大学学術情報リポジトリ
Keio Associated Repository of Academic resouces

Title	コンテクストネットワーキングを実現するCatalogue system
Sub Title	
Author	金子, 晋丈(Kaneko, Kunitake)
Publisher	慶應義塾大学デジタルメディア・ コンテンツ統合研究センター
Publication year	2018
Jtitle	慶應義塾大学DMC紀要 (DMC review Keio University). Vol.5, No.1 (2018. 3) ,p.16- 26
JaLC DOI	
Abstract	
Notes	特集 DMC研究センターシンポジウム第7回 デジタル知の文化的普及と深化に向けて： コンテクストネットワーキングの分散型ミュージアムへの展開 開催日時：2017年11月24日 (金) 14:00～17:30 開催場所：慶應義塾大学日吉キャンパス西別館1 講演録
Genre	Departmental Bulletin Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO32002001-00000005-0016

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

コンテクストネットワーキングを実現する Catalogue System

金子 晋丈
（慶應義塾大学理工学部専任講師・
DMC 研究センター研究員）

今日のタイトルは、コンテクストネットワーキングとカタログシステムということでお話をしたいと思います。

まずコンテクストネットワーキングについてです。コンテクストというものについて先ほど渡部先生の方からとても分かりやすい発表をいただいて、何も私からは説明を付け加えることが何もないと思いながら拝聴しておりました。ですが、まず、コンテクストネットワーキングについて私なりの理解をまとめさせていただきまして、その後コンテンツをネットワーク化すること、そしてそのコンテンツをネットワーク化してできたコンテンツネットワークをどのように設計していけばいいのか、技術的な話をさせていただこうと思っております。私たちは実際にコンテンツネットワークとしてカタログシステムというものを作成しており、本日は、カタログシステムの構成とそれがもたらす世界、また、これからの技術的な挑戦についてお話しできればと思っております。

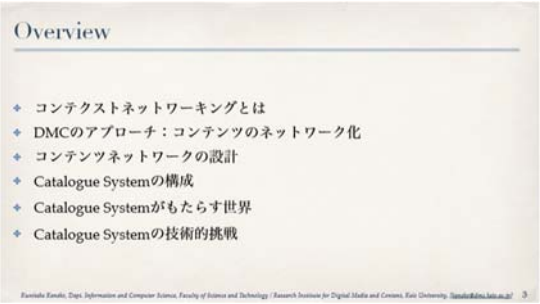
先ほど、きれいなゴッホの絵や自画像を見た後ですが、福沢諭吉先生の胸像を例に話します。まず、コンテンツは基本的には無数のコンテクストの中に存在しているというのが私たちの考えている立ち位置です。

例えばここに福沢諭吉の胸像があります。これは三田キャンパスの中にある胸像を写真撮影したものです。これをコンテクストで見ると、三田キャンパスにあるので三田キャンパスの彫刻の中にも位置付けられますし、福沢諭吉ということであれば 1 万円札や、福沢諭吉がアメリカで撮った写真、咸臨丸の話、また、この胸像は誰が作ったか、誰が撮った写真なのか、といったさまざまなコンテクストをこの 1 枚のデジタルファイルに与えることができます。

このつながりをどのように使っていくか、ということが、コンテクストネットワーキングになるわけですが、例えば福沢諭吉の胸像をスタートポイントにして、このいろいろな矢印で表されるコンテクストを見渡し、あらゆるコンテクストを知ることで、このコンテンツに対する理解が進みます。

これはまさに先ほど渡部先生がおっしゃっていた内容だと思います。また、1 つのコンテクストをたどり、選択し、先に進んでいくことによって、別のコンテンツの発見につながる。これを繰り返し実現していくのがコンテクストネットワーキングであると

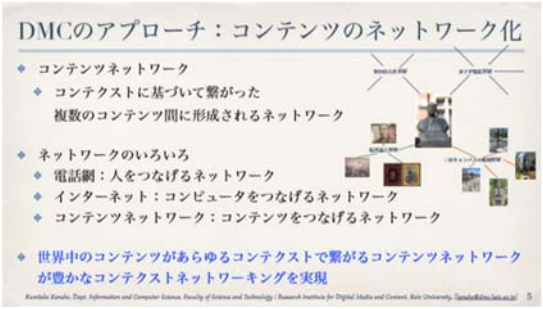
考えております。



DMC はこのコンテクストネットワーキングの実践に当たり、コンテンツをコンテクストでネットワーク化することがそれを解決する技術的な手法だと考えております。すなわち、ここで先ほど出てきたコンテンツをコンテクストでうまくつなげてあげることによって、コンテンツ間をつなげるネットワークをつくってあげようということになります。

従って、コンテンツとコンテンツをつなげるのはコンテクストであり、もっと広い視野で見てみると、複数のコンテンツ間にさまざまなコンテクストでネットワークが構成されていくことになります。皆さんは今、コンテクストとコンテンツというものを 1 個のペアとしてとらえているかもしれませんが、それを重層的に組み合わせることで、コンテンツネットワークを目指そうとしているのです。

ネットワークというと皆様の頭の中に浮かぶのは、例えば電話網、もしくはインターネット、コンピューターをつなげるネットワークが頭に浮かんでくると思います。我々がやろうとしているのは、コンテンツとコンテンツをつなげるネットワークを新たにつくっていくことです。それをコンテンツネットワークと呼んでいるわけです。



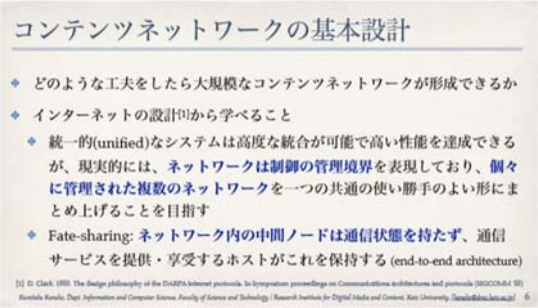
では、どのようなコンテンツネットワークをつくれば、コンテクストネットワーキングの世界を最大限に楽しむことができるのでしょうか。それは、コンテンツの数を増やす、コンテクストの数を増やすということにほかなりません。

なぜならコンテクストが少ない、コンテンツが少ないと「ああ、そうだよね」と、完結しやすいためです。そこに新しい発見は何も得られませんし、新しい理解も得られません。

従って、いろいろな人がいろいろな価値観を提示する。それを同じ場所に持ち込むことによって、コンテクストネットワーキングは非常に豊かになっていくと考えることができます。

そうしますと、技術的には、いかに大規模なコンテンツネットワークを構成していくかが肝になってくるというのは明白であろうと思います。ここがコンテンツネットワ

ークを技術的に構成していくための、基本設計のスタートポイントにあたるわけです。



コンテンツネットワークという言葉自体が皆様にとっては初めてかもしれません。多くの人にとって初めてだろうと思います。さらにそれを大規模にするというのは、どのようなことなんだろうと、思われるでしょう。ここで過去の事例から、大規模なネットワークをつくるために何をしなくてはいけないのかといったことについて話してみたいと思います。

私がご紹介する事例は、インターネットの設計です。皆さんはおそらくインターネットを当たり前のように使い、その設計がどういうものか、どういう考え方にもとづいてインターネットというものがこの世に存在することになったのか、そういったことについてはあまりご存じないかもしれません。

THE DESIGN PHILOSOPHY OF THE DARPA INTERNET PROTOCOLS

David D. Clark

Massachusetts Institute of Technology
Laboratory for Computer Science
Cambridge, Ma. 02139

Abstract

I. Introduction

そこで本日は、『The design philosophy of the DARPA Internet protocols』という論文を持参しました。

IT系の方々は聞いたことがあることかもしれません。これはデービッド・クラークという MIT の計算機科学者が、1988 年に発表した論文です。内容は、約 15 年前に TCP/IP が開発されて今に至るが、実はその技術的な背景が誤解されていたため、ここでまとめておこうと思う、といった主旨の論文です。

ページ数はそれほど多くないものの、非常に深みのある、今また読み返してみても面白い論文です。私は、大学生のときに、この論文と出会いました。

大学生のときに読んで、分からないなりに納得しましたし、今日の講演に向けて久しぶりに読み返してみたところ、味わい深く、さまざまな言葉が心に突き刺さり楽しい思いをしました。この論文は、インターネットのデザインの中で、大規模にネットワークをつなげていくためには、何をしなくてはいけないのかといったことを示唆しています。

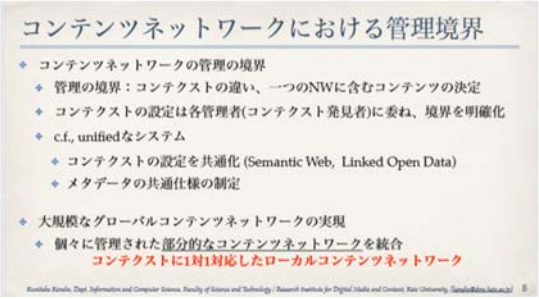
ゴールとして、まずファンダメンタルなゴールと、セカンドレベルのゴールの 2 つに分けています。ファンダメンタルなゴールの中にはこのような表現があります。『ユニファイドなシステムは高度な統合が可能で高い性能を達成できるが、(インターネットは) 現実的には、ネットワークは制御の管理境界を表現しており、個々に管理された複数のネットワークを 1 つの共通の使い勝手のよい形にまとめ上げることを目指す。これによって大規模なネットワーク空間をつくる』ということを言っております。

すなわち大規模なネットワークを津々浦々まで 1 つの方式で統一し、これをみん

なでやるんだというような方法を取るのではないのです。個々個別の組織が持つ状況やネットワークに求められる機能、それはそのまま維持した状況で、差異がある状態でそれらをつなげていく。各ネットワークの違いを了承した状況でつなげていき、つなげたときに１つの共通の使い勝手のよい形(ア・コモンユーティリティー)を実現するということがファンダメンタルなゴールとして挙げられております。

もう１つ、セカンドレベルのゴールから抜き出してきたのは、フェートシェアリングという考え方です。これはエンド・ツー・エンドのアーキテクチャーとインターネットが言われるもとになった考え方です。インターネットの基本的な考え方では、クライアント端末とサーバーが通信に関する状態を、どこまでデータを送ったか/届いたかといった状態として管理しサービスを提供しています。インターネットの真ん中にはルーターやいろいろなスイッチがあるのですが、そこは一切のサービスに関する通信状態を持たない。端末が壊れたらサービス終了。でもこのクライアントとサーバーが生きているのであればサービスは継続すべきである。それ以外のほかのところが状態を持たないようにしよう、という考えです。こういった内容が、実はインターネットの設計に書かれています。私はこれがインターネットをここまで拡大させることになった、とても重要な設計指針であると考えております。

さて、先ほどやや難しい言葉が出てきました。制御の管理境界ということですが、制御の管理境界を簡単に言うと、LAN です。自分の LAN、家の LAN、ホームネットワーク。家のブロードバンドルーターがあって、プリンターとかコンピューターをつなぐと思いますが、それは自分自身で管理できます、ほかの人は管理しません、という考え方です。



では、この考え方をコンテンツネットワークに適用した場合はどのようになるのでしょうか。我々はコンテクストネットワーキングをコンテンツネットワークで実現しようとしていますので、コンテクストの違いが管理の境界になるのではないでしょうか。それはほかの人が踏み込めるものではなく、その人自身が管理するととらえることができます。

別の表現をすると、LAN と同等のローカルコンテンツネットワークに存在するコンテンツは自分で決定することができるということです。それが、管理の境界が明確ということかと思います。

すなわち、コンテクストの設定とは、コンテンツネットワークにおいて、各管理者、いわゆるコンテクストを発見した人に委ねて境界を明確化することが、先ほどのインターネットの設計指針から得られる、大規模ネットワークを構築するための重要な策ではないかと考えるわけです。

ここで参考までに、デービッド・クラークの論文において、大規模なネットワークを

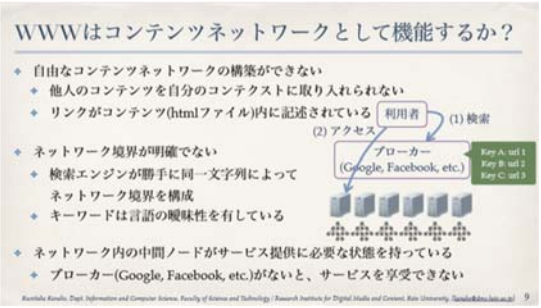
つくるまでのオルタナティブアプローチとして出てきているユニファイドなシステムとコンテンツネットワークの関係についてお話しします。ユニファイドなコンテンツネットワークシステムでは、あるコンテクストをこのように定義し、その定義にマッチするものを、このように整理しなさい、というような一種のルール決めを行い、それに完全に従うことになるかと思います。

具体的な例を挙げると、メタデータの共通基準をつくり、それにもとづいてデータを管理しましょうという考え方は、ユニファイドなシステムを大規模につくろうとしているにほかならないと思うわけです。

また、セマンティックウェブ、もしくはセマンティックウェブから派生して出てきたリンクドオープンデータは、自由な記述が特徴ですが、一種の検索エンジンのような中央集中型のものがウェブ空間上クロールしなければその情報にたどり着くことはできないのです。

すなわち大規模で、グローバルなコンテンツネットワークを形成しようとするときに、どのようなアプローチをとるかと言いますと、個々に管理された部分的なコンテンツネットワークをひとつひとつ構築することになります。部分的なというのは、コンテクストに 1 対 1 対応したローカルなコンテンツネットワークを積み重ねていき、大規模なグローバルなコンテンツネットワークをつくるということです。これが、大規模ネットワークをつくるための鍵になってくると考えられます。

さて、最近ウェブサーフィンやネットサーフィンといった言葉をまったく聞かなくなりましたが、ウェブはコンテンツにアクセスできるメカニズムです。では、ウェブはコンテンツネットワークとして機能するかを考えてみたいと思います。



まず、HTML ファイルについてです。ウェブというものは HTML ファイルがあり、そこにリンク情報があります。始点となるコンテンツにアクセス権がないと、つまり、自分が編集できる状態でないとリンクを記述できないという問題があります。すなわち HTML ファイルに書くとリンクを張れるが、HTML ではないファイルから、他の誰かのウェブページへのリンクを加えたいと思っても加えられない。すると自由なコンテンツネットワークが実現できません。

次は、ネットワーク境界が明確ではないことが挙げられます。具体的に言うと、皆さん検索エンジンを使っていると思いますが、検索エンジンというのは同一な文字列を用いて、その文字列でグループ化をしているにすぎません。その文字列がネットワーク境界を構成しているわけで、私たち人間が選別して構成しているわけではありません。さらに、文字を使っているというところが大きな課題となっております。例えば、キーワードは言語でできているため、文字を使った曖昧性が生じるわけです。入力した人、もしくはアクセスする人も、その言語の使い方に違いがあるとうまくネットワーク境

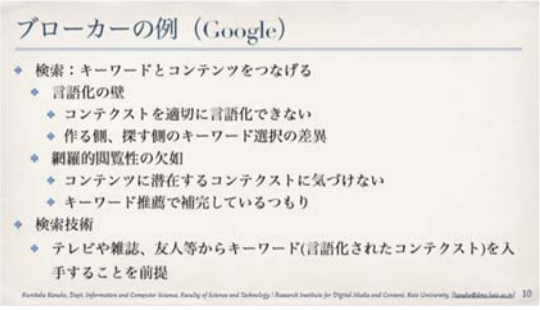
界を定めることができないといえます。

最後に、ネットワーク内の中間ノードがサービス提供に必要な状態を持っているということがインターネット、WWW のもう1つの大きな課題になっていると思います。

実際に例を書いてみました。利用者＝ウェブブラウザーです。下にウェブサーバーがあり、「Google」、「Facebook」、私はこれらを勝手にブローカーと呼びました。この例から、利用者はまずブローカーに行き検索をして得られた結果でアクセスをすることがインターネットアクセスの現在の構図になっています。ブローカーは何を持っているのかというと、キーワード A を含む URL は何々ですよという対応付けのリストを持っています。

そのリストをなくしてサーバーにはアクセスできないということが、ウェブサービスの課題です。すなわち、ブローカーがこれを変えることによって違うところにナビゲートされてしまう。本当は欲しいデータはここにあるのに、テーブルが異なっていると他の情報源に行かざるを得ないということが WWW では発生します。

これがデービッド・クラークの書いた論文の 2 個目のネットワークの中のノードが状態を持っているというところに反している例と考えることができます。



ブローカーについて、具体的に「Google」の例と「Facebook」の例を出してみました。「Google」は、何と何をつなげているコンテンツネットワークかといいますと、キーワード（文字列）がコンテンツをつなげています。先ほどお話したように、キーワードを使っている以上言語の壁というのがそこには立ちはだかります。コンテクストというものは、適切に言語化できない限りそれをキーワードで探し出すことはできません。

そして、作る側、探す側でキーワード選択の差異が発生すると、文字列で識別しているので、その差異を埋めることができないというのが言語識別の壁です。

さらに、網羅的閲覧性の欠如がキーワード検索にはあります。キーワード検索はキーワードを入れるとコンテンツが出てくる検索です。コンテンツは複数のコンテクストを持っていますが、それらのうち自分の知らないコンテクストは、それに見合ったキーワードを入れない限り、永遠に出てこないというのが網羅的な閲覧性の欠如です。

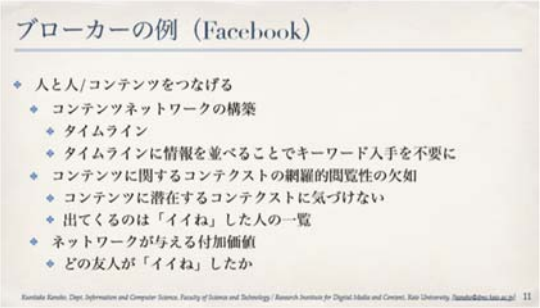
「Google」はある日から、キーワード推薦を実装してきました。キーワードを 1 つ入れる、2 つ入れると、候補となる 3 つ目のキーワードが列挙されてくるというものです。

これは、一種、網羅的閲覧性の欠如を回避するための技術と言えるかもしれませんが、一方でそれはよく使われるキーワードを順番に出しているに過ぎないため、結果的には存在しているコンテクストを埋もれさせてしまうシステムになっているのかもしれません。

一般に「Google」のような検索技術は、最初にどこからか、コンテクストが与えら

れなければならず、これに対して、何も手助けをしてくれません。テレビや雑誌、友人等からキーワードが与えられる前提のサービスです。

すなわち言語化されたコンテクストを誰かから入手しないと、「Google」検索サービスというのは使えないことになります。



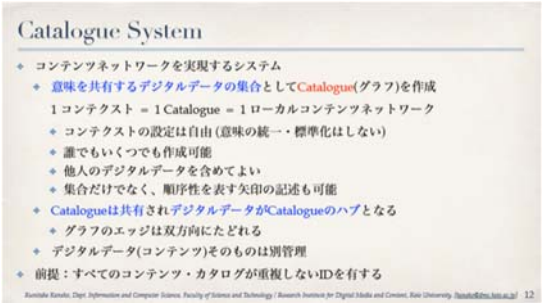
一方、「Facebook」は、一種のコンテンツネットワークであると言えます。

これは人と人、もしくは人とコンテンツをつなぐコンテンツネットワークでしょう。非常に面白い仕組みを持っており、コンテンツネットワークはタイムライン、つまり時系列で並ぶという特徴があります。タイムラインに情報を並べることで選択が可能になり、従って、言語にしなくていいといったコンテクストの言語化を省いていると言ってもいいかもしれません。

一方で、コンテンツに関するコンテクストの網羅的な閲覧性ができているかというと、推薦されたコンテンツは上がってきますが、そのコンテンツにどんなコンテクストがあるのかを提示してはくれません。

このコンテンツに誰々さんがいいね！しましたという情報が出てくるだけで、基本が人です。ソーシャルネットワークと言われるように人を中心としたネットワークなので、いいね！した人の一覧が出てきます。

ネットワークを考えるときにその付加価値についても考えますが、これは先ほどのデービッド・クラークから得られる示唆とはやや離れますが、どの友人がいいね！しているから見よう、といった行動が「Facebook」の大きな売りになっております。その付加価値はここにとどまっているということになります。



さて、我々が考えているコンテンツネットワークを実際に具現化したものが、このカタログシステムというものです。カタログシステムはコンテンツネットワークを実現しているもので、意味を共有するデジタルデータの集合としてカタログを作成します。

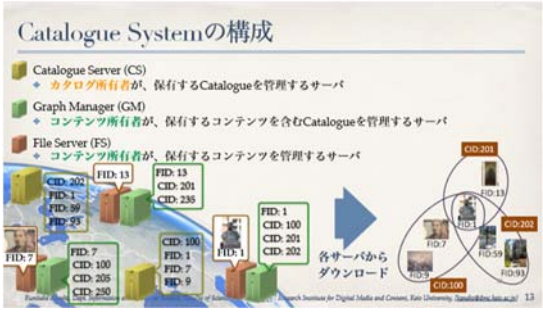
すなわち、カタログシステムでは、1コンテクストが1カタログで、1つのコンテクストを共有するコンテンツ群が1つのカタログの中に入ってきます。それで1つのローカルコンテンツネットワークを形成するという考え方です。

誰でもいくつでもローカルコンテンツネットワークを作成可能で、他人のデジタルデータをこのカタログの中に入れておき、そのカタログの中で順序を示すような記号を加えても構いませんよ、というのが、カタログの考え方です。

そして、このカタログをグローバルに共

有する。共有してどのように使うのかといいますと、このカタログの中に入っているデジタルデータが、複数のカタログのハブになります。1 つのデジタルデータが複数のカタログのハブになることで、コンテンツネットワークを形成することがカタログシステムの特徴になっています。

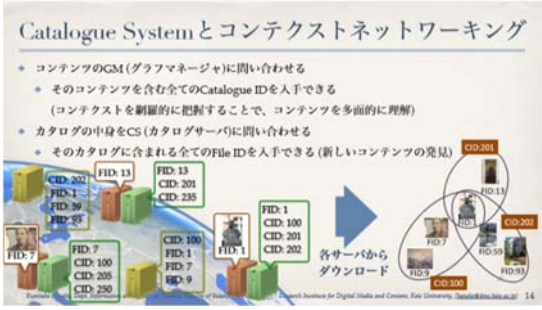
カタログシステムでは、このようなつながりだけを管理するので、コンテンツそのものは別管理したり、デジタルデータをカタログのハブにするため、すべてのコンテンツカタログが重複しない ID を有することが前提のシステムとなっています。



さらに具体的にシステムの中身を見てみますと、カタログシステムは大きく 3 つのサーバーで構成されます。カタログサーバー、グラフマネージャー、ファイルサーバーです。皆様にとって馴染み深いものはおそらくファイルサーバーです。

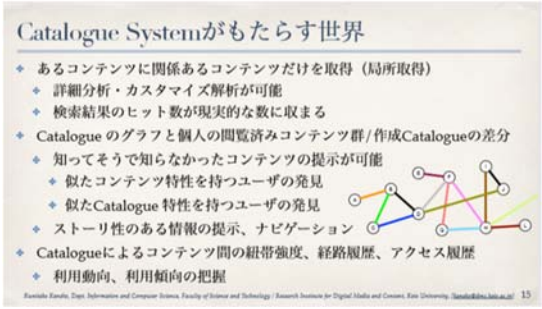
先ほどカタログを発見した人、つまり新しいコンテクストを発見した人は、自分のカタログサーバーに保有するカタログを保存していきます。

グラフマネージャーは、コンテンツ所有者が持つサーバーで、先ほどのハブを実現します。デジタルデータがハブで、そのデータにどのようにカタログがスポークとして存在するかを示すサーバーです。



この仕組を図示しますと上記のようになります。ユーザー側ですべて取得し、組み上げてみるとこのようなコンテンツネットワークが出来上がります。

さて、この仕組みを使ってコンテクストネットワーキングがどのように実現されるのかと言いますと、このサーバーから問い合わせてきて出来上がったここを見たときに、ファイル ID 1 番のこのコンテンツに対して、カタログが 3 つある、つまり 3 つのコンテクストが登録されているということが分かります。それを実現するのがグラフマネージャーになります。



1つのカタログに注目するということは、つまり 1 つのコンテクストに注目することになります。そして、別のファイルを発見するのが、カタログサーバーの果たす役割になります。

このようなカタログシステムがどのような世界をもたらすかといいますと、可能性としては上記のようなことが挙げられます。

ここに書いてある内容は、とても奥深いものが並んでおりまして、簡単に説明できるようなものではありませんが、代表的なものとしては、1 つのあるコンテンツに関係のあるコンテンツを高速に取得する、局所取得が可能になります。

「Google」が行っているような、ウェブ空間を全部クロールしてデータをかき集めてくる方式に比べてみますと、自分に関係するデータだけを持ってきて、そこの部分だけで詳細の分析、カスタマイズされた解析が可能になります。

一方、現在の「Google」は例えばキーワードを入れた結果、数百万件の検索結果が出てきます。皆さん、あの膨大な量の検索結果は嬉しいものでしょうか。おそらくは、全く嬉しくないですよね。きっと最初に出てきた 5 個程度しか見ないと思います。つまり数百万件の結果を出す価値はありません。

検索結果のヒット数が、現実的な数に常に収まっている必要があるとするならば、ある程度の絞り込みができるシステムが、良いシステムなのではないかと思います。

次に、パーソナライズにつながっていくという観点です。カタログからグラフ構造を持ってきて、その中の何個を自分が知っているか、自分がすでにアクセスしたかというものを比較することによって、自分が見ていないコンテンツを浮き彫りにする。その関係あるコンテンツで、見ていないコンテンツを浮き彫りにすることができるのがこのパーソナライズ、グラフを使ったパーソナライズのサービスです。

ただ単に閲覧したコンテンツを見るというだけではなくて、閲覧したコンテンツを軸にグローバルのコンテンツネットワークとの差分を取ったり、自分がコンテンツ間に引いているグラフを確認することで、自分の物事の理解ができるようになります。グラフは、いわゆる自分が発見したコンテクストを表しているので、サーバー側にある誰かのカタログとその構造が似ていれば、同じものの理解をしている可能性が高いということになります。すると、似ているユーザーを発見することもできますから、また新しいサービスが可能になるのではないかと思います。

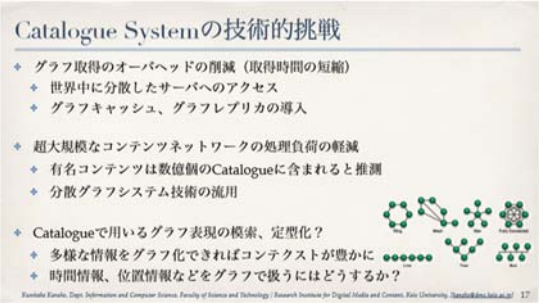
そして、グラフ構造、コンテンツのカタログで構成されるグラフは、図のようになりますが、あるコンテンツがどの程度強い関係を持っているのか、ということも計算できるようになっていきます。

DMC ではこのカタログシステムを用いて、ナビゲーションシステムを 2 つ作っております。1 つはモザイクのビジュアライズです。もう 1 つは、キャンパスミュージアムというウェブサービスです。こちらはコンテンツネットワークのグラフの形、トポロジーと私たちは言っておりますが、トポロジーを表現して可視化し、それをクリックすることで新しいものを発見していく。トポロジーを俯瞰的に見ながらサービスをするアプリケーションです。

キャンパスミュージアムは、バックエン

ドにこの構造を持っているのですが、可視化はせずに、例えばこのコンテンツを軸に複数のカタログを提示するようなウェブサービスに仕立て上げております。

最後になりますが、カタログシステムの技術的挑戦をいくつかご紹介したいと思います。



先ほどよりお伝えしておりますとおり、コンテンツネットワークといっても、1 つのネットワークがあるのではなく、コンテンツとそれをつなぐリンクのようなイメージです。グラフの話だとエッジと呼びますが、エッジは、世界中のサーバーにばらばらと保存されています。いろいろなサーバーから情報をかき集めてこないとコンテンツネットワークは作ることができません。

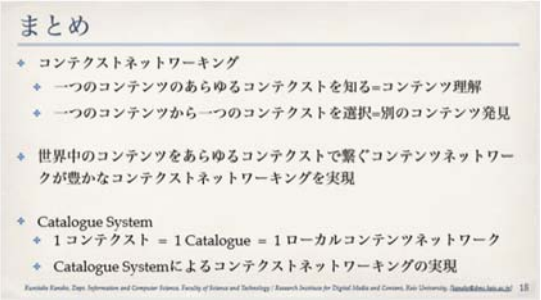
これを作るためには、グラフの情報を世界中からかき集めてこなければならないのですが、そのオーバーヘッドはかなり大きなものが必要になります。

これを解決するために、グラフキャッシュやグラフレプリカというものを設置し、よりユーザーの近くで、もしくは少ないクエリーでその情報を取得することを研究としています。

また、超大規模なコンテンツネットワークの処理負荷が課題です。例えば、ルーブルのモナリザといった著名なコンテンツは、皆さんが関連付けをしたがります。そのような作品になりますと、規模として数兆のカタログをどのように取得するのか、それをどのように処理するのか、という課題が生じます。

逆に、コンテクストを把握するといった際には、ルーブルのモナリザなどは恐らく数億のコンテクストとつながっているため、実は全てのコンテクストを把握することはもうほぼ不可能に近いということになります。

次に、グラフ表現をどうするのかという課題です。今は、時間情報、位置情報などもグラフで扱い、それをコンテクストとして表すことができればもっと面白いカタログシステムになるのではないかと思っております。このあたりはまだまだ挑戦しなくてはいけない分野と思っております。



最後に、コンテンツネットワークというのは、1 つのコンテンツが持つあらゆるコンテクストを知ること。コンテクストを知ることでコンテンツを理解しようという試みです。そして、あるコンテンツから、1 つのコンテクストを選択することで、別のコンテンツを発見しようという試みがコンテクストネットワーキングだと考えています。これらを実現するためのシステムとしてカタログシステムを作っています。

本日、インターネットの設計指針からコンテンツネットワークを作り、維持していくためには、何をしなくてはいけないのか。それらを踏まえた上で、現在はカタログシステムを設計しているということをご説明させていただきました。ありがとうございます。

金子 晋丈（かねこ くにたけ）
慶應義塾大学理工学部専任講師・DMC 研究センター研究員。専門はアプリケーション指南ネットワーキング。特に、デジタルデータの利活用を促すデジタルデータのネットワーク化について研究を行っていいる。2001 年東京大学卒業。2006 年同大学院情報理工学系研究科博士課程終了、博士（情報理工学）。同大学院新領域創成科学研究科での特任助手を経て、2006 年 9 月より慶應義塾大学デジタルメディア・コンテンツ総合研究機構、特別研究助手。2007 年、同機構特別研究講師。2012 年 4 月より現職、デジタルメディア・コンテンツ統合研究センター研究員を兼任。