

情報爆発のこれまでとこれから

Info-plosion : Retrospection and Outlook

喜連川 優

Abstract

情報爆発なる現象に二つの視点から論ずる。第1は情報過多な時代において情報に溺れることのないための情報検索技術の課題とそれに対する取組みについて、第2は過多とも思える情報を逆手にとって利活用する技術について述べる。とりわけ後者に関しては新しい研究開発が活発になりつつある。情報爆発から派生する様々な課題を研究すべく2004年に文部科学省科学研究費特定領域に申請する段階では、サイバーフィジカル、あるいは、smarter planet という提唱はなかったが、今日、「情報爆発からの価値創出」に関して多様な試みがなされるようになった。人、モノ、事象等、社会システムにおけるあらゆるものがIT的に捕捉可能な時代、すなわち、超可観測の時代に突入し、新しいサービスが躍動的に生まれようとしている動きを解説する。

キーワード：情報爆発，サイバーフィジカルシステム，情報検索

1. はじめに：how much information?

人類の活動はITにより一層多様化し、その創出する情報量は爆発的に増大しつつある。とりわけその増大は21世紀に入り著しく、IDCは2020年に35ゼタバイトにも達すると予想している⁽¹⁾。図1に過去から将来における人類が創出する情報量の増大の様子を示す。そもそも情報量の推定は容易ではないが、発端としてUCバークレイ校のHow much informationなるプロジェクトの取組みが広く注目された⁽²⁾。これは1999年に対する解析を2000年に発表したもので、2003年にも再度試みられている。その後は継続的にIDCがその予想を公開情報として発表してきている。情報分野における研究課題は多岐にわたるが、情報爆発という現象は情報分野全体に対して非常に大きなインパクトを与えるものの一つと考え、2004年秋に筆者らは文部科学省科学研究費特定領域研究に「情報爆発時代に向けた新しいIT基盤に関する研究」という申請を行った⁽³⁾。Jim Grayが、大量の情報サイエンスのあり方を大きく変えている傾向、すなわち、スパコン中心の計算科学からデータセントリックな科学へのシフトの重要性を第4のパラダイムと

して指摘したことは広く知られているところである^{(注1), (4)}。2011年1月NSF長官Dr. Subra Sureshはその講演で、explosion of informationとそれを支える情報基盤の重要性を指摘しており⁽⁵⁾、今日においても、情報爆発という現象が与える変革に関して多くの研究者が注視しているところといえよう。本稿では筆者が深く関わった文部科学省の「情報爆発特定領域研究」と経済産業省の「情報大航海プロジェクト」での成果をベースに種々のこれまでの動きについて振り返ってみるとともに、今後の潮流について考えたい。なお、執筆時点ではBig Dataなる更に新しい言葉が生み出されつつある⁽⁶⁾。

2. 情報爆発に対する視点

我々は、情報爆発という現象には二つの側面があると考えた。Wikipediaにもinformation explosionなるエントリが2001年に作成されているが、overloadというネガティブな側面がとりわけ指摘されている⁽⁷⁾。情報が増えれば増えるほど、欲する情報の全体に対する割合は小さくなり、探すことがどんどん大変になり時間を要する作業となると予見することは自然な見方であり、実

喜連川 優 正員：フェロー 東京大学生産技術研究所
Masaru KITSUREGAWA, Fellow (Institute of Industrial Science, The University of Tokyo, Tokyo, 153-8505 Japan).
電子情報通信学会誌 Vol.94 No.8 pp.662-666 2011年8月
©電子情報通信学会 2011

(注1) スパコンの出力は極めて膨大であり、第4のパラダイムは第3のパラダイムの中核であるスパコン計算結果の活用基盤の重要性を指摘するものであって、スパコンのサイエンスにおける重要性は当然のことながら不変である。

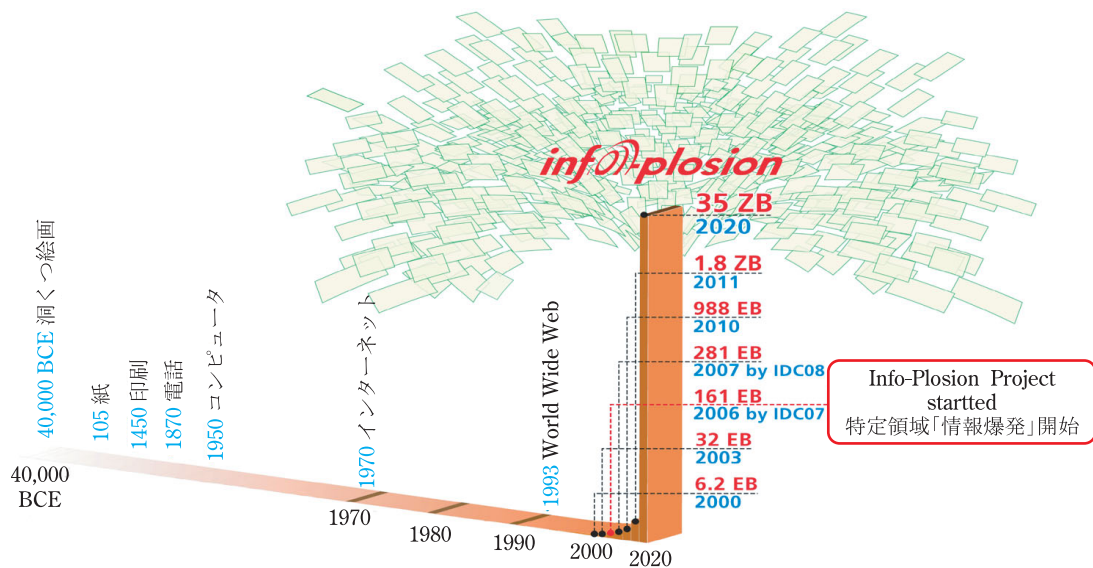


図1 情報爆発⁽¹⁰⁾ (出典: Horison Information Strategies, cited from Storage New Game New Rules, p. 34 (www.horison.com), IDC, The Diverse and Exploding Digital Universe 2020 (http://www.emc.com/collateral/demos/microstites/idc-digital-universe/iview.htm))

際、最近ではググ男という言葉さえ生まれてきている。検索はもはやファッションな作業ではあり得ず、ちょっと複雑な検索をしようとする現行の検索エンジンでは歯が立たず、何度も検索語を入れ替えるなど人が強く介入する必要があり、面倒なので他人にやってもらおうという発想である。もちろん、Webと検索エンジンによって人類の生活は飛躍的に利便性が向上したことは事実であるが、情報量の増大の割合が、検索技術の向上を凌駕しているともいえ、とりわけ、現在の検索技術では、ナビゲーション的な問合せはおおむね十分な能力を有するものの、やや複雑な質問をしようすると、検索窓を前にして途方に暮れることも多く、情報過多な時代をネガティブに捉える意見が多いのもうなずけるところである。検索、情報を探すという一見ありふれた古典的な行為は、まだまだ先の長い研究が必要といえる。

一方、情報爆発に対するポジティブな視点とは、これほど大量の情報に人類が遭遇することはこれまでの歴史にはなく、これを新しいチャンスと捉えることもできるのではないかという視点である。大量のデジタル化されたデータ、つまり、コンピュータが操作可能な対象がこれだけ膨大に生まれたのは今世紀であり、先端的な多くの情報技術を駆使することにより今まで見つけられなかった「価値のしずく」を絞り出せるのではないかと考えることもできよう。

膨大なデータが生成されるようになると、そのデータの調理手法としての情報技術を新たに生み出す大きな駆動力が伴ってきた。データマイニング技術の多様な展開はその一例といえよう。e-Bayは2009年に1日に50ペタバイトのデータを処理していると発表した。リーマンショックの後にもその売上を伸ばしたのは、膨大なデー

タを処理し購買傾向を精緻に解析可能な先端的なIT基盤によるところが大きいと推察される。また、スマートグリッドが精力的に模索されているが、米国では、その効率的運用のためのセンサーデータの規模は1年分でエクサバイトに達するという報告がある。大量データはうまく活用すれば、大きな価値に転換できる、あるいは、従来にない全く新しいIT領域がこの大量データの解析分野から生まれるという見方である。

情報爆発という名称からはどちらかというと前者のネガティブな側面が第一印象として捉えられがちであったが、筆者らは、前者の解決のみならず、後者のポジティブな側面の模索に関しても4.で述べるごとく大きく注力してきた。

3. 検 索

情報爆発に対するネガティブな側面を解決することの一步は欲しい情報を容易に見いだすことのできる検索技術の高度化である。Webという膨大なかつ自由に利用可能なテキストが利用可能になったことにより、「言語資源」が著しく豊かになり、言語処理環境は大きく変容したといえる。この資源を利用することにより、言語処理手法そのものの高度化が加速され好循環を生むと2004年当時想定した。図2に情報爆発特定領域研究における黒橋らの結果を示す(横軸は解析テキスト量、縦軸が精度を表す)。また、ページランクはナビゲーションクエリに対して有効ではあったものの、皆が知っていないマイノリティとしての情報や意見の価値も重要であり、その抽出は依然課題として残る。最終的には精緻な言語処理を適応する時期が到来するはずであり、該領

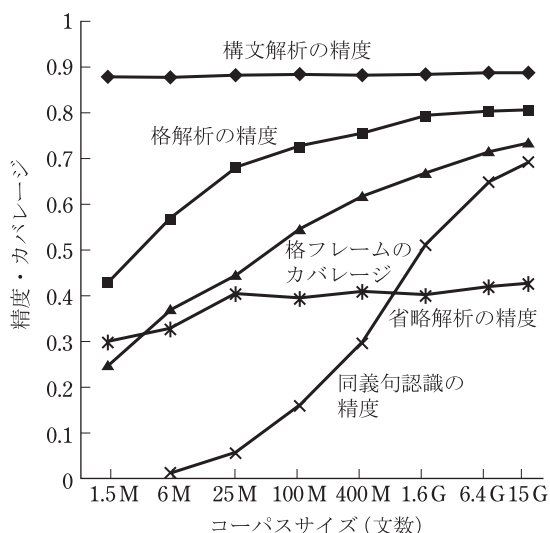


図2 解析テキスト量と検索精度

域の基礎研究が不可欠と考えた⁽⁸⁾。2010年、IBMはgame changing serviceという言葉を入れているが、同じ競争の場で戦うのでは利得が少なく、むしろ根底から視点を変えることが不可欠と考えた。1秒でURLリストを返す現行の検索エンジンは、広告モデルに大きく依存しており、少々時間がかかってもよいので欲しい情報をきっちりとまとめてくれる分析エンジンを使いたいユーザも多いのではないかと考えた。先行者利益が極度に大きいインターネットビジネスでは、ちょっとした差別化ではユーザの流動は見込めない。2011年IBMはワトソンなる知識エンジンを数年がかりで開発し、米国の歴史あるクイズ番組で優勝した。チェスのdeep blueに次ぐ素晴らしい快挙といえよう。現状で利用可能な応用は多々あると想定されるが、固有有名を中心とした簡潔な回答の検索に限定されているという点では、まだまだ研究の余地があるともいえる。加えて、wikipedia, dbpediaなど人間によって知を編集したサイトの存在がなければ、この勝利はなかったと開発に携わった研究者が報告しているが、自然言語でやりとりし難題に対し人間よりも正しく答えられるという実証は検索技術の次の姿を示唆したといえる。

情報爆発特定領域研究においては、事実の検索に限定せず人間に一步でも近づける難しい質問に答える分析エンジンを目指して5年間の基礎研究を進めてきた。このような活動に共鳴した多くの言語処理研究者が協力して、とりわけ、黒橋(京大)が中心となりTSUBAKIなるオープンプラットフォームを構築してきた。鳥澤(北陸先端大, 現NICT)による600万もの上位下位関係ペアも組み込まれている。例えば、「パソコンで肩が重くならない姿勢」というような質問は、通常の検索エンジンでは答えられないものの、TSUBAKIでは答えることができるに至っている。「風邪に効く野菜」という質問

をしても、上位下位関係辞書を利用し具体的な野菜名を答えることができる。検索サービスは様々な要素技術の複合体であり、大学の研究者が個人プレーで個々の要素技術を開発するだけでは、その有効性をユーザに問うことは極めて困難といえる。大規模な情報資源を対象とする研究では、互いの技術を持ち寄りそのシナジーを実現可能なプラットフォームが不可欠であると考えた次第であり、今後も有効に機能すると考えている。紙面の都合上詳細の紹介はできないが、例えば、藤井(東工大)による震災復興税、赤ちゃんポスト等賛否両論で容易に正しいとは結論を出せないトピクスについての賛成、反対の意見をまとめるopinion reader等、多様なサーチを実現し、更に、ユーザ評価も実施した。2004年当時、検索窓一つでどこまで質問者の意図を探ることができるかが課題であった。現時点でも同様の研究が底流にあるものの、一方で、バーティカルなサーチ、すなわちドメイン特化によるシャープな検索が重要であることも強く認識されるに至っている。上記のように種々のサーチ機構に挑戦しユーザ評価も行い、おおむね好評を得てきた。

検索を高度化するにつれ、分析・解析との境界が次第にはっきりとしない状況になりつつある。New York Timesが2006年にperson of the yearをYOUとし、「貴方が世界を変える主役である」というメッセージを発信したことは非常に印象深かったが、今日、6億を超えるユーザを抱えるfacebookや、対人口比では日本人のユーザが最大であるTwitterをはじめとするソーシャルメディアの存在感が大きく向上しており、それに連動し、当該メディアの解析がとりわけ企業におけるマーケティングの観点から精力的に進められている。Twitterはリアルタイムメディアとして従来の放送を越える大きな存在感を明らかにしてきた。テキスト情報源は古典的なWebから多様なITメディアへと展開しつつあり、その爆発的大量なUGC(User-Generated Contents)の解析、とりわけリアルタイム解析は今後ますます情報爆発資源の中心的な活用スタイルとなってゆくものと考えられる。

4. サイバーフィジカルな世界

Googleは世界中の情報をorganizeするとその社是としてうたったこともあり、情報爆発は検索エンジンによって解決されるだろうと考えた人が少なからずいたと想定する。実際2008年のGoogle blogには1兆ページの捕捉が報告され、その数に圧倒されたのも事実である。ただ、同年に平均ページサイズやディープWebの研究報告などもなされており、クロウラの対象となるいわゆる表層Webは、IDCが予想する情報全体空間からすれば極めて僅かではない。内容の濃さは別として、量という観点では、所詮テキスト情報は僅かではない

のは想像に難くない。サーベイランスカメラやデジタルシャドー等を含めた多様なセンサによる情報量が膨れ上がることは確実といえる。筆者らは、これからの情報爆発の源泉はここにあると考える。

情報爆発特定領域研究では、いわゆるメタボ検診を対象としたヘルスケアに関して、運動による糖尿病など生活習慣病の予防を目的として加速度センサを用いた運動量管理システムを中島（九大）、須藤（東大）を中心に開発してきた。2008年には経済産業省情報大航海プロジェクトにおいて「情報薬」と命名し、規模感のある実証実験を行いその有効性を確認した。抗生物質のように菌を殺傷する強力な薬ではないものの、生活習慣病ではちょっとした示唆（情報）を与え、後押しするだけで、運動不足が解消できることが多いことを実証した。副作用の精査など今後も継続的に研究すべき課題は数多く残されているが、通常の薬の開発コストに比べると情報薬のそれは十分小さいと想定される。該システムの実現には、個人からの膨大なセンサ情報の収集並びに該情報から運動種別を特定する解析システムが肝となる。まさに、情報爆発の積極的活用といえよう^{(9),(10)}。

情報大航海プロジェクトでは、デジタルヘルパーなる時空間アラートサービスも構築した。これは日常の行動パターンを携帯電話のGPSを活用することにより捕捉し、その行動範囲と移動パターンにパーソナライズした情報推薦を行うサービスである。アマゾンの本の推薦が極めて有効であることは有名であるが、あくまでもサイバー空間内での購買履歴情報のみを利用したものである。それに対して、実世界の活動パターンを利用することにより、はるかに個人に適合した情報提供が可能になるであろうとの考えから開発が模索されたサービスであるが、実際、アクセプタビリティは高く、また、コンバージェンスも飛躍的に大きく向上することが3,000人規模の実証実験で明らかになった。この場合にも、個人の行動を示すGPSセンサ情報や購買履歴等多様な情報は膨大となり、それを解析する基盤が肝となる。同様に情報爆発の積極的な活用といえよう。なお、東京海上日動火災は、このプロジェクトの成果を活用しドコモとワンタイム保険を製品化した。

IBMは2007年からsmarter planetというビジネスを展開している⁽¹¹⁾。2050年には人口の70%が都市に住むと予測し、とりわけ、smart cityなる都市インフラ全体の高度化をITによって実現することを狙っている。その制御対象は、水質、エネルギー、交通など多岐にわたり、多様な物理量を精細に観測すると同時に、高度な解析により社会インフラの大幅な効率化を実現可能としている。グーグルがサイバー世界の情報を一手に集約し、そこから多様なサービスを生み出しているのに対し、実世界からの多様なデータを収集し、高度なサービスを生み出そうとしており、好対照ともみなせる（もちろん

グーグルも最近ではスマートグリッド等にも進出しており、検索語のマーケットに次ぐサービス市場を見据えていることは明らかである）。smart cityは観測対象物により広く捉えているものの、先の事例同様、肝は、実世界の観測から発生するセンサ情報爆発とその解析にある。電力使用のみならず多種多様なセンサが提案されている。同様の考えはシーメンスが古くから行っており、また最近、多くの企業が参入し始めている。HP社はCentral Nervous System for the Earth (CeNSE)と命名し、同様の研究開発と製品化を進めている。

米国NSFは2006年にサイバーフィジカルシステムなる研究に着手している⁽¹²⁾。実世界とサイバー世界のより密な連携により実世界をより高度化することを目指しており、smarter planetの提唱以前に既にそのほう芽を見ることができる。スマートグリッド、ヘルスケア、航空等多様なアプリケーションを当初から想定し、それぞれの分野で数多くのワークショップを開催してきているが、それだけではなく、システム構築のための基礎についても注力している。2011年3月に開催された情報爆発最終シンポジウムにおいて、その活動が報告されたが、年間3,000万ドルの予算が継続的に投入される予定とのことであった。ソフトウェアの開発という視点に立つと、今日の多くの機器は既にコンピュータを内蔵しているものがほとんどであり、その機器特性は組込みソフトウェアによって形成されているといえるが、一般的に組込みのCPUは非力である場合が多く、高度な最適化は困難である。サイバー世界においてより強力な解析パワーを活用することにより、システムの高度化が期待される。組込みとサイバー世界のソフトウェアの連動フレームワークは新たな戦略ソフトウェアとなることが期待され、そこが一つの注力点である。従来のプログラミングモデルは抽象度と場所の透明度を上げ、ユーザには物理世界を意識させない高度化が追求されてきたといえる。サイバーフィジカルシステムでは逆に、物理世界を意識したソフトウェアモデルが必須となる。また、都市のような大規模なシステムのみならず、体内システム等ミクロな世界も対象とする等、NSFは多様な基礎研究を推進している。

このほかにも、多くの試みがある。例えば米国では2007年にエネルギー自立安全保障法においてNet-Zero Energy Commercial Buildings (ZEB) Initiativeを規定し、2030年までに米国に新築される全ての業務用ビルのZEB化を推進している。東大では野城らが新しいZEB教育棟を建設中である。膨大なセンサ情報の解析を行うことにより精緻な暖冷房制御を行い段階的ネットゼロ化を実現する予定である⁽¹³⁾。

筆者らは、最先端研究開発支援プログラムにおいて、次世代の戦略的社会的サービスの構築を図3に示す形で実現することを描いている。近年開発が著しく進められて

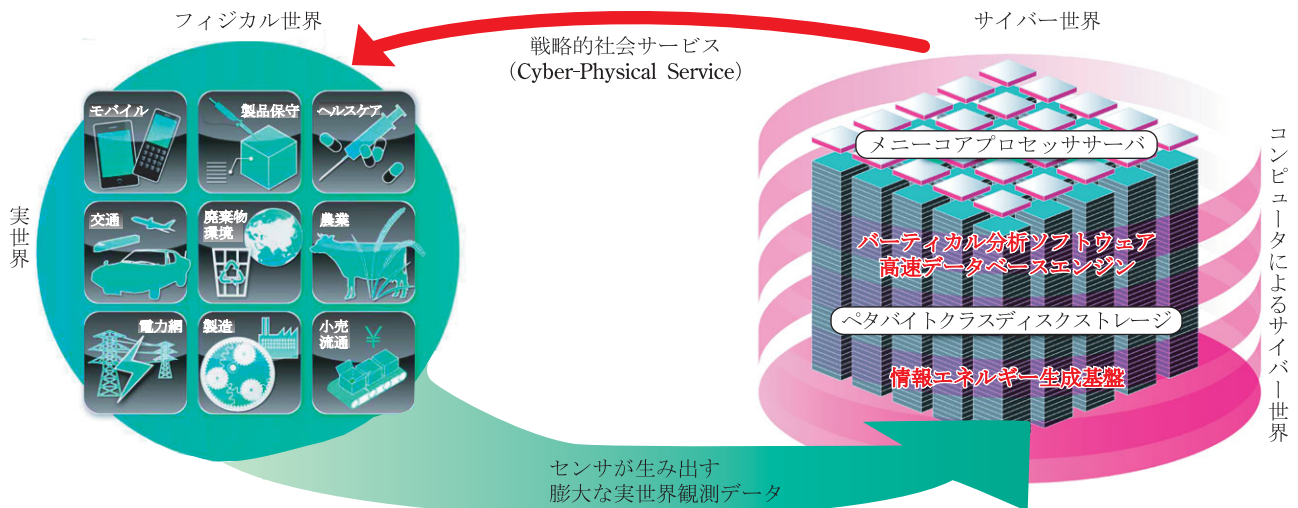


図3 最先端研究開発支援プログラムにおける Info-energy Generator 概念図

きた MapReduce, Hadoop, Dryad 等に見られる大規模データ処理系はいずれも強力な CPU 演算パワーを要求していない。すなわち、浮動小数点演算能力を追求したいいわゆるスーパーコンピュータとは大きく異なり、大容量データの入出力帯域に力点を置いたデータインテンシブアーキテクチャが、いわゆるインターネットメガサービス企業において威力を発揮している。大規模データを高効率に操作するデータベースエンジンが重要となるとの考えから情報爆発指向の新しいエンジンの開発を進めている (図3)。

5. おわりに

情報爆発なるプロジェクトを提案した当時、地球上の人間は有限であり、情報の生成はすぐに飽和するのではないかという意見もあったが、今日、価値あるデータを生み出す主役は人間のつぶやきではなく、センサや人工物となりつつあり、情報爆発は一層加速するものと想定される。

情報爆発を克服する検索技術、情報爆発を活用するサイバーフィジカル技術という二つの観点で技術の動きを眺めてみた。筆者の認識不足も多々あるやもしれず、是非御批判・御指摘頂ければ幸いである。

上記の二つのいずれの視点においても情報爆発の「料理」の基盤としてのいわゆるクラウドの果たす役割は大きい。今回の震災において、地震が影響するエリアのクラウドへの被害は極めて少ないと聞く。環境変動に強固なデータセンターを集中的に構築することの有効性が明らかになった。電力面でも柔軟な制御が可能であり、適度な分散と資源集約を追求するクラウドとの併存は、今後の姿といえよう。情報爆発プロジェクトではマイクロソフト社より大量のクラウド資源の無償提供を受け、Windows Azure としては過去最大規模の実験を行い若

手研究者は貴重な経験を得ることができた⁽¹⁴⁾。

今後ますます主流となる情報爆発指向のデータインテンシブな研究開発を推進する上で、研究者が存分に利用できる大規模クラウド環境の整備が我が国においても不可欠と考える。

文 献

- (1) http://gigaom.files.wordpress.com/2010/05/2010-digital-universe-view_5-4-10.pdf
- (2) <http://www2.sims.berkeley.edu/research/projects/how-much-info/>
- (3) <http://www.infoplosion.nii.ac.jp/info-plosion/>
- (4) <http://research.microsoft.com/en-us/collaboration/fourthparadigm/>
- (5) http://www.nsf.gov/news/speeches/suresh/11/ss110113_scidplo.jsp
- (6) http://en.wikipedia.org/wiki/Big_data
- (7) http://en.wikipedia.org/wiki/Information_explosion
- (8) 喜連川 優, 松岡 聡, 松山隆司, 須藤 修, 安達 淳, “情報爆発時代に向けた新しい IT 基盤技術の研究,” 人工知能誌, vol. 22, no. 2, pp. 209-214, 2007.
- (9) “特集 情報爆発時代におけるわくわくする IT の創出を目指して,” 情報処理, vol. 49, no. 8, pp. 879-955, 2008.
- (10) “特定領域成果概要集「Info-plosion—情報爆発—情報爆発時代に向けた新しい IT 基盤技術の研究」,” 平成 23 年 3 月 8 日発行, 2011.
- (11) <http://www.ibm.com/smarterplanet/us/en/>
- (12) http://en.wikipedia.org/wiki/Cyber-physical_system
- (13) <http://www.iis.u-tokyo.ac.jp/publication/topics/2011/20110304press1.pdf>
- (14) http://cloud.watch.impress.co.jp/docs/news/20101001_397473.html

(平成 23 年 4 月 22 日受付 平成 23 年 5 月 19 日最終受付)



喜連川 優 (正員: フェロー)

東大教授, 生産技術研究所戦略情報融合国際研究センター長, 東大地球観測データ統融合連携研究機構長, 1983 東大大学院工学系研究科情報工学専攻博士課程了。工博。本会副会長, IEEE Int'l. Conf. on Data Engineering Steering Committee Chair, 文部科学省科研費特定領域研究「情報爆発」領域代表, 経済産業省「情報大航海」戦略会議委員長。本会業績賞受賞。データベース工学の研究に従事。