最先端研究開発支援プログラム（ＦＩＲＳＴ）平成２２年度フォローアップに係るヒアリング（超巨大データベース時代に向けた最高速データベースエンジンの開発と当該エンジンを核とする戦略的社会サービスの実証・評価）

１．日時　　平成２３年９月２８日（水）１６：３０～１７：００

２．場所　　中央合同庁舎４号館１２階　共用１２０２会議室

３．出席者

相澤　益男　　総合科学技術会議議員

本庶　　佑　　総合科学技術会議議員

奥村　直樹　　総合科学技術会議議員

川本　憲一　　政策統括官（科学技術政策・イノベーション担当）付参事官（最先端研究開発支援プログラム担当）

４．説明者

喜連川　優　　東京大学生産技術研究所教授（中心研究者）

安達　　淳　　国立情報学研究所教授（研究支援統括者）

上田　修功　　国立情報学研究所客員教授

５．議事

【川本参事官】

これより研究課題「超巨大データベース時代に向けた最高速データベースエンジンの開発と当該エンジンを核とする戦略的社会サービスの実証・評価」の平成22年度フォローアップに係るヒアリングを始めさせていただきたいと思います。

本日の総合科学技術会議側の出席者についてはお手元の座席表のとおりです。

このヒアリングについては非公開で行います。また、関係者がフォローアップを通じて知り得た情報は、フォローアップの目的のみに使用させていただきます。ただし、後日、今後の研究発表あるいは知的財産権等に支障が生じないことを確認させていただいた上で、議事概要を

公開させていただきます。本課題では研究課題側からの配付資料のうち、回収資料と明示されたものについては、ヒアリング終了時に回収をいたしますので、あらかじめご了解いただきたいと思います。３枚紙のものであります。

時間配分につきましては研究課題側からのご説明を10分、その後、質疑応答を20分、合計30分の時間厳守でよろしくお願いします。説明に当たりましては、終了３分前に予鈴、終了時間に本鈴を鳴らさせていただきます。時間が来ましたら、質疑応答を優先するということで説明が途中であっても、そこで中断をお願いしたいと思います。質疑応答につきましては終了３分前に予鈴を鳴らさせていただきます。それではご説明をよろしくお願いします。

【説明者】

それでは、東京大学の喜連川を中心研究者とする私どもの研究プロジェクトをご説明いたします。私、研究支援統括者の安達がご説明させていただきます。

私どもは新しい実行原理に基づく高速データベースエンジンの開発と、それによる新しい社会サービスの創出のための基盤技術の開発に取り組んでおります。２つのサブテーマがございまして、データベースエンジンに関するものは喜連川が、そして新しい社会サービス創出のための基盤技術開発に関しては上田が担当しております。

この図は私どもが考える巨大データ時代の次世代IT基盤の図でございます。センサーネットワーク技術が大変進歩しまして、膨大なデータが生み出されてきております。この膨大なデータによって、実世界の様子を精緻に観測することが可能となってきました。私どもは「情報エネルギー生成基盤」と名付けておりますが、いわばクラウドの中にこの膨大なデータを集積し、そこで大容量のデータを高度に解析することによって、サイバーの世界から物理世界へフィードバックするという形に今後のITシステムが発展していくと考えております。

このサイバーの世界から物理の世界への大きなサイクルが、従来の組み込みシステムを基本としたITとは異なり、ダイナミックに変化する社会システムを対象とする今後のITの姿であると私どもは考えております。これを実現するためには、大規模データを管理し処理するための高速データベースエンジン、ならびに多様かつ斬新な社会サービス創出のための分析ソフトウェア、この２つの要素の開発が極めて重要と考えた次第です。

まず、第１のサブテーマである高速データベースエンジンの開発をご説明いたします。図の左側が従来型の順序型の実行原理によるものですが、中心研究者はこれに対して図の右側にあ

ります非順序型という新しい実行原理を提案しました。非決定的な順序での処理を可能とし、大量の非同期入出力を発行することにより、非常に高速なデータベース処理を実現することがこのエンジンの特徴です。この図の左側の従来型に対して、新しく提案する方式では図の右側のように密度の高い入出力処理をすることによって、大幅な性能向上を目指せることがお判りいただけるかと存じます。

このための研究環境としまして、ここに示します図のようなシステムを導入致しました。サーバー、大規模なストレージ、それをつなぐスイッチから構成されるものでありますが、残念なことに震災後の電力の逼迫によりまして、実際にはこの図に示しますような小規模な構成で実験を行って参りましました。

さて、その研究の内容でございますが、東大におきましては非順序型実行という基礎的な研究ならびにオープンソースのデータベースソフトウェアを使ってのアルゴリズムの実装を、また、これと並行しまして、パートナー企業の日立製作所では、我が国で唯一自社開発の関係データベースソフトHiRDBに対し、東大が創案しました非順序型アルゴリズムを導入して、抜本的に作り変えるというソフトウェア開発を進めて参りました。現在までに性能評価実験により100倍以上の性能が出ることを確認することができております。

ソフトウェアの動作は分かりにくいため、可視化して、ビデオでご覧に入れたいと思います。図の右側の赤い点が新しい方式、左側の青い点が旧来の方式で、横軸が時間を表し、縦軸は上図ではIO帯域を、下図ではアドレス空間を表しております。中心研究者の創案しました非順序型方式では非常に密度の高い処理を実現可能でありますことから、すぐに処理が終わってしまいます。一方、旧来の方式ではまだまだ続いておりまして、おおよそ100倍ぐらい時間がかかっているということがわかります。

この辺でビデオを止めましてスライドでの説明を続けたいと思います。今年６月に東大生産技術研究所の一般公開がありまして、そこで東大と日立がプレス発表を行いました。これは日立が製品化を来年に行うというもので大きな反響を生みました。本来の計画ですと、再来年度に全機能を実装し製品化する予定でありましたが、これを早めまして、一部の機能でも早くリリースした方が、インパクトが大きいと考えこのようにしたものです。

さて、サブテーマ２の社会サービス創出のための基盤技術の開発に移りたいと思います。このスライドはアメリカの現在の様子をまとめたものです。昨年暮れにアメリカの大統領への答申であるDigital Futureが出されました。５月にはマッキンゼーがビッグデータ、巨大データ

に関するレポートを出しております。また、８月にはCyber Physical Systemというテーマで NSFが会議を持ちました。いずれもトッププライオリティをヘルスケアに置いております。それ以外にもいろいろな社会システムの課題はあるわけですが、アメリカにおいては、トッププライオリティはヘルスケアと位置付けられており、医療費削減のために10年以内に病院の４分の１の治療を在宅に移そうという目標等を掲げております。

私どもはもちろん医療費のこともございますが、生活習慣病などの場合、日常生活における予防医療が不可欠であるとの考えから、これを一つの課題として設定し具体的な実証実験のテーマとしております。生活習慣病の様々な患者の情報をセンサーネットワークよりモニタリングするとともに、その莫大な情報を解析し適切な生活指針の情報を提供するというサービスを研究開発しております。

さて、そうは申しましても最初から在宅治療の実証実験をするというのはなかなか困難ですので、まずは第一段階として私どもは済生会熊本病院の協力を得まして、病院の入院患者で模擬的にこの実証実験を進めております。生活習慣病の患者に対しまして血圧を始め様々なデータをモニタリングするわけですが、加えて、今後は在宅で介護士がさまざまな介助サービスをする際の行動とその有効性をモニタリングすることも重要と考えられ、看護師の行動情報もモニタリングしております。このように医療行為と看護行為のモニタまでご協力頂けることは大変珍しく、従来なかった新しい実証研究に挑戦致しております。

患者の膨大なデータ、ならびに看護師の行動データの分析を行うためのマイニング技術につきまして研究を進め、転移学習に基づく新しい手法を提案し、有効性を明らかに致しました。この成果に関し著名な国際会議でも学術的に高く評価されています。このように社会サービス創出のための基盤技術の開発に一歩を踏み出すことが出来ましたが、今後、一層深化させていく所存でございます。

最後になりましたが、最近、ツイッターと呼ばれるマイクロブログメディアがよく使われるに到っておりますが、それが震災当日にどのように有効に機能したかにつきまして、膨大な情報の分析例としてデモでご紹介致したく存じます。

【説明者】

このデモンストレーションでは、震災直後において、ツイッター上で人々が一体、どのように情報を伝播しているかというような様子を可視化してお見せします。ツイッターから15億以

上の莫大な数のツイートを収集しておりまして、その分析を行うソフトウェアを作成致しました。

これが、震災が起きた直後のサイバー空間ですが、ここから時間を進めていきますと、ご覧いただけますように様々な情報が塊となって広がっていく様子が分かります。この大きな塊は阪神大震災の経験者が、まず第一に、風呂に水をためなさい、次に、ガスの元栓を閉めなさい等のアドバイスを経験に基づいてツイートしておりまして、それがこの図の上では線、ツイッターではフォロワーというリンクがあるのですが、これでつながった色々なユーザーに対して、極めて短時間のうちに大変多くの人々に伝わっている様子がご覧いただけます。

更に時間を先に進めていきますと、非常に様々な情報が数時間のうちに大量に共有されていく様子がこの様に見てとれます。これは19時頃のサイバー空間の状況ですが、幾つかの伝搬情報の例をご紹介させて頂きますと、この塊ではビックカメラで充電器を無料で提供する情報がツイッターで広められたことが、あるいは、自販機が無料開放されたことの情報伝搬などを確認出来ます。また、避難情報、避難場所をグーグルマップスというグーグルの地図サービスの上に整理して公開された方がおられ、非常に便利で、私も利用しましたが、非常に多くの人に伝搬し、共有されていったことがわかります。

このように、この度の東日本大震災においては、震災後の数時間で、非常にさまざまな有益な情報がサーバー空間上の多くのユーザー間で共有されるということが起きておりまして、大量のツイート情報を収集・分析することにより、その拡散プロセスを解明できることをご覧いただきました。以上で報告を終了いたします。私どもの研究へのご支援に心より感謝致します。

【川本参事官】

ありがとうございました。それでは、これより質疑応答のほうに移りたいと思います。ここからの進行については、相澤先生、よろしくお願いします。

【相澤議員】

進捗状況を伺わせていただきました。それで、非常な勢いで研究開発が進んでいることは理解できているのですが、基本的なところでお伺いしたいのですが、センサーワールドとコンピュータワールド、これをつなげるというところが基本ですよね。それで、サイバーの世界をセンサーを介してリアルワールドとつけると、そういうことなのですが、センサーの方のインプ

ットはどういうような形でこの研究開発に関連して進んでいくのか、これは既にあるものをただつなげていくというコンセプトなのかどうか、そこのところをまずお伺いしたいのですが。

【説明者】

大変有益なご指摘をありがとうございます。米国が今日のITをどう見ているかということを説明したスライドの図を見ていただくと一番わかりやすいと思うのですが、ちょっと小さくて恐縮ですけれども、現代をAge of Observationという言い方をしております。即ち、今まではコンピュータの上に組み込みソフトウェアが載って個々バラバラに動いているだけの状態でしたが、今日におきましてはセンサーテクノロジーがネットワーク化され、それがサイバー空間にコネクトされ、世の中の全体が可観測になった。これが非常に大きな変革であると見做しております。

（研究推進上支障が出るおそれがあるため事例説明は非公開）

Cyber Physical Systemと米国において名付けられた考え方は、実社会の状況を常に捕捉し、同時に過去のストック情報も一緒に解析し、全体を精緻に解析することにより、従来と比べて大幅な効率化と、変化への迅速な対応を可能な社会システムの構築が可能と考えております。ポイントは、サイバー空間で多様な情報を広く集約し、解析する基盤を構築しようという次世代IT感にあります。色々な分野でこのような手法が有効と考えられておりまして、サイバーとフィジカルをくっつけるということが今後の大きな流れであると考えております。これは、実は米国がサイバーフィジカルという名称を生み出した2006年より以前に、私共は2004年より情報爆発IT基盤と呼んでいたものと合致します。

【相澤議員】

そこで、結局、最高速度のデータベースのエンジンだと言われているわけですが、センサーワールドの方もそれに伴って大容量のインプットが入ってくる、こういうところに対応するには、これだけの性能のエンジンを作らなければいけないという、そこが見えないと、これでどういう意義が出てくるのかというところが分かりにくいです。というのは、先ほど例としてどこかの医療機関とのこの例を出されていますよね。これだとイメージとして今までのセンサーがただ単にぱらぱらあると、こういうようなものを対象にしておられる。何でこんなに高速の

ものが必要かということになりますので、その辺を。

【説明者】

ご指摘のとおりです。医療関連のサービス実証に関しましては、患者の同意をとらなくてはいけない、医師の同意もとらなくてはいけない、倫理委員会も通さなくてはいけないというような多くのハードルがございまして、やはりその規模は、奥村先生にもご指導いただきました情報大航海プロジェクトでもそうでしたが、かなり小さなものにならざるを得ないのは事実です。しかしながら、長い目で見ますと、在宅介護が1千万人規模となることを想定致しますと、その時のビッグデータというものを見据えて、今、ここで着手しなくてはならないと考えております。

（研究推進上支障が出るおそれがあるため事例説明は非公開）

【相澤議員】

それで、そういう理解ですと、この研究課題としての具体的なターゲットは、次のステップはどこなのかというのがもう少し分かる形で示していただきたいことと、それから色々な実際の適用例、何が対象かによる適用例を次々と開発していかれるようにも見えるのです。でも、それだとすると、これもできる、あれもできるということをただ適用のシステムを広げていくだけのようにも見えます。そうでありますが、これだというはっきりと分かるところに絞っていくという方式はとられないかどうかということ。

【説明者】

まさに最先端研究ということで、そういうターゲット自体を模索しながら研究を進めようとしております。今年、マッキンゼーが出したレポートでも、ビッグデータが世界を変えていくだろうとレポートしておりますが、その中心メッセージは「物がしゃべる時代になる」ということです。IoTというような表現がされることもあります。センサーがしゃべるような時代になったときに、ディスラプションが出てくるだろうとレポートしておりますが、それがどこで革新的に花が開くか、これははっきり申しまして、まだ誰も分っていないのが現状であると思います。

そんな中で、やっぱり、現状のペインのレベルをぐっと落とせる領域というのはどこかとい

うのを幾つか模索している状況ではないかと感じておりますが、私どもはそれにチャレンジしながら共通基盤技術を絞り出し構築してゆくという、地味な作業かもしれないですが、学術的にはそのようなアプローチが妥当と考えております。しかし、現状では、先生にご指摘頂きましたように、一つ一つバラバラにやっているだけに見えてしまい統一的な基盤技術というものまで深化させた姿をご提示できておりません。まだ１年少々ですので、今後、最終年度までにはしっかりとしたものを構築したいと考えております。すなわち、幾つかのバーチカルなものが立ったとしても、バーチカルというのは先生がおっしゃっている個別という意味ですけれども、個別の中から要素的な共通軸というものがやはり出てくるだろうと思います。そういうことも含めまして、実はアメリカのNSFと一緒に随分密に連携をしながら、おたくはこれをやる、うちはこれをやるというようなことから、知見を共有しようということを今、本当に積極的に進めております。４月にも米国に行きましたし、８月にも行きまして、向こうとの意思疎通を高めております。例えば一つにはこういう大きなデータを扱うということから、そこをどのように圧縮するのか、しかも圧縮をしながらその上で機械学習ができるのか、つまり、圧縮を解いてからやるのではなく、圧縮したものの上でできるかというようなデータの加工の課題もあります。さらに推論エンジンにつきましても、現在までの研究状況を振り返りますと、例えば教師なし音声認識の発展には30年ぐらいかかっているわけです。

今日われわれは、音声のシグナルに代わってセンサーのシグナルを扱おうとしているわけですが、今から次の10年で課題解決ができるのかは誰も分からないのが実情です。しかし、そのような技術が世の中を大きく変化することだけは確実です。そのための要素技術を我々はコツコツと作りながら、例えば在宅医療では、太った体格の人でも、非常にスキニーな人間でも、そこそこ扱えるようなある種の共通の運動認識技術というようなコア技術はどんどん出せると考えております。

今までは、映像、音声そしてテキストと、メディアというのはこの３つしかございませんでした。しかし、言わば、第４のメディアとしてセンサーストリームというものが登場したと考えております。それに対しての基盤技術を作ることを目指しております。アプリを横展開できる基盤を作っていこうと考えております。ただ、そのためには、今の段階では、最初から一般化を狙うよりも、まず、バーチカルな個別の問題からアタックしていく、世界的に見ても、そのような状況になっていると認識しております。

【相澤議員】

ですから、そういう状況が分かりますし、これはやはりプロジェクトですから、このプロジェクトの今、おっしゃったようなことを戦略として示していただけると明確だと思うのです。そうすれば、このプロジェクトとしては、ここまで達するのが当面の大きな目標で、明確になってくるのではないか。少し心配いたしますのは、先ほどのように来年、もしここでまたヒアリングをやったときに、こういう例もまた出てきましたというような報告になることはないのか。ただ、それだけだと先ほど言いましたような心配もありますので、メーンのストラテジーにはこういうのがあり、そこのためにはこういうことの要素をきちんと積み上げてとか、そのようなところの戦略性が見えるとよろしいのではないか。

【説明者】

本当に先生のおっしゃるとおりで、先生から賜りました貴重なご意見に対して、全く反論するつもりは毛頭ございません。ただ、コンピュータサイエンスは使われるようになりまして50年が経ちまして、非常に難しい時代に入ったということは事実だと思います。画像処理を例に挙げますと、画像処理のパッケージを作ろうとするときに、トンネルの中でコンクリートが落ちてくるのを防ごうとするひび割れの画像処理と、半導体の欠損を検出するための画像処理、これらは同じアルゴリズムで済むのか、という問題があります。やっぱり、それぞれ個別に対応する部分を相当に入れて対処するという時代に入ってきていると思います。ですから、そういう意味で、ここでのバーチカルというのは、実際に役に立つアプリを構築するという観点ではある程度はしかたがない時代に入ってきているというのもご理解いただきまして、その厚みのある部分が何なのかというのを今、必死になって我々は検討しているとご理解いただけるとありがたく存じます。

【奥村議員】

今の話と関係するのですが、バーチカルにどれだけクリティカルなものをやはり対象として選ばれてやるかで決まると思います。これは必要だと私も思います、バーチカルは。だから、今回の例が余り病院の例は少し私もいかがかと、こういう感じが私はするのですけれども、私の質問は、これはたしか700倍か800倍にすることが目標だったと思うのです。要するにその意味とも関係する。まず、その前に100倍まできた。あと、仮に800倍だと８倍速くすると、課題

は何ですかと、ほとんどないのでは。そんな言い方をすると何でしょうかということが一つと、それから、２つ目はそもそも非順序型の実行原理のこのコンセプトは、何か権利的に守られているのでしょうかというのがもう一つ、プラクティカルな質問です。

それから、もう一つ、質問しますと、何となくもちろん、これはうまくいく対象と、そうでないケースがあると思うのですが、せっかく、ここまでできているのに何かこれをもってITの一つのベンチャーでも立ち上げるというようなお話は全くないということは、一体、何なのだろうかと。３つ。

【説明者】

答えやすい部分から答えさせていただきますと、２番目の権利化に関してですが、実はこの最先端の提案をさせていただいた段階では、もちろん、特許になっていますという前提では全然ない状態でございました。2005年に我々はパテントを出したわけですが、日本に関しましては、昨年、成立いたしました。

【奥村議員】

日本では成立しましたか。

【説明者】

はい。米国に対しましても、現在、進めております。

それから、何でベンチャーを起こさないかという話は、私が25年前にも似たような話を頂戴しました。アメリカ人から「おまえのこの論文のアイディアで、何でベンチャーを起こさないのか」と何度か言われたことがございますやはり、先生、日本の中でベンチャーを起こすことの難しさというのは、非常に大きいのでないかという気がいたしまして、私はむしろ大学の役割というものは、ある種、研究を突き詰めるというところではないかと思うこともございます。また、開発の部分に関しましては、現時点でデータベースのコードの規模OSのコードより大きいぐらいです。そのコードを維持することは、今のベンチャーの技量でやるのは、非常に大変になっているのではないかと感じている次第でございます。一方、私どもとしましては、日本を元気にするような成果を出すことが最先端研究開発支援プログラムの重要な役割であると感じておりまして、そういう意味で、我が国で唯一、自身でデータベースのソフト開発を行って

いる日立さんと協力させていただいているということでございます。

さらに、先生の最初のご質問の800倍についてですが、最初の100倍は大変かもしれないが、100倍まできていたら残りの８倍ぐらいはそれほど難しくないだろうと、そうお考えになられるかもしれませんが、性能向上率を上げれば上げるほど困難性が高くなりまして、先ほどパワーポイントでお見せしました機械の構成図ですが、800倍の高性能化を実証するには相応の巨大なシステムが必要となって参りますが、まず、あのシステムがまともに動作するかということを検証するために半年以上かかっています。

このように、環境構築自身が非常に大変でございます。システムが落ちますとわけのわからない挙動というのがたくさん出てまいりまして、それを一個一個つぶしているのにも時間がかかっております。そういう意味で、余り直球のお答えになっていないかもしれないですが、現状版の100倍と言いましても限定された試験項目に対して、その可能性を示したということでして、製品化という視点で見ますと、まだまだ実現できていない部分が多く残っているのが実情でございまして、現在、懸命に開発を急いでいる状況でございます。

【奥村議員】

それであればこそ、先ほど申し上げたように性能に見合う適切な対象をやはり選ばないと、プロジェクト全体としてはバランスのとれた姿に仕上がりにくいですね。

【説明者】

（研究推進上支障が出るおそれがあるため本説明は非公開）

【本庶議員】

一言、非常に期待しているのですが、医療が非常に重要だというのはまさにそのとおりなので、これから実はゲノムコホートをやろうと思っていて、電子情報とゲノム情報とあらゆるビヘービアも含めてやっていきますと大変なことなので、それを解析するソフトウェア、システムは非常に大きな課題。アメリカもそれを結局、認識して、そういうことを言っているので、是非とも、そこのところには注目して、電子カルテはまだちょろいので、電子カルテが何百と集まって、つまり、各病院のサーバーからたくさんデータを集めてきて、研究者がそれをまた患者さんの他のデータと合わせて、それで何か意味を出す。さらに、そこにDNAの情報が入っ

てくると大変な事業です。ですが、これは必ず将来、この問題は良いソフトでやらないとできないと思っていますので、是非、そこのところは常に、先生、フォローしていただいて、まだ、今、そこまではデータ自身もいっていませんが、やがて、その時代はすぐに来ると思いますから。

【説明者】

本当に先生がおっしゃっていただいた通りで、そのようなことをやろうと思っております。とにかくデータが巨大なものですから、巨大なものから情報を集める機能は、誰がやろうとしましても絶対に必要です。そこをまず一つ作ろうと言うことをやっております。解析の部分はドメインの方とお話をしなければいけません。私どもが一番やりたかったのは、実は厚生労働省が先日出しましたレセプトとメタボのデータで、これを実はやりたくてしようがない次第です。ご配慮賜れますと幸甚でございます。我々としてはぜひ解析基盤を構築できればと考えております。

【本庶議員】

レセプトデータは使えることになったでしょう。

【説明者】

現実には、容易に扱える状況でもないと伺っております。是非お手伝いさせて頂ければと考えております。

【川本参事官】

これでヒアリングを終了させていただきたいと思います。ありがとうございました。

－了－