

3D Stroller : 三次元情報空間散策システム

津高 新一郎 有田 英一 安井 照昌

RWCP* 新機能三菱研究室

* Real World Computing Partnership

概要

我々は新しい情報検索システムのプロトタイプである 3D Stroller を現在開発中である。このシステムでは、文書及び単語は各々ユーザが操作可能なオブジェクトとして表現され、各オブジェクトは互いの関連の深さが距離の近さに反映されるように三次元仮想空間内に自己組織的に配置される。ユーザは自分自身の興味や観点に応じてオブジェクトの移動や削除等の操作を行うことが可能であり、ユーザの操作に伴い各オブジェクトの再配置が行なわれる。このような仮想空間の再構成を通して検索式が自動的に生成され、関連の浅い文書が三次元空間から自動的に削除される一方、関連の深い文書が検索され三次元空間に随時登録される。本システムは大量の文書を対象にユーザの興味や観点に基づく情報空間の構築とその中での散策を可能にする。

3D Stroller : the System for Strolling in the 3-Dimensional Information Space

Shin-ichiro Tsudaka Hidekazu Arita Terumasa Yasui

RWCP* Novel Functions Mitsubishi Laboratory

* Real World Computing Partnership

Abstract

We are now developing a '3D Stroller', a prototype of a new information retrieval system. In this system, all the documents and words are represented by objects which can be manipulated by users, and each object is placed in the 3-dimensional virtual space by self-organization, according to the relationship with other objects. Users can move or delete these objects according to their preference or viewpoints. It is followed by the reallocation of other related objects by the system. Though these procedures, search queries are automatically generated, and related documents will appear in the 3-dimensional virtual space, while unrelated documents will disappear. This system enables users to construct the information space and stroll in it, based on the users' preference and viewpoints in large document databases.

1 はじめに

近年における社会の情報化、ネットワーク化は個人が得ることのできる情報量を飛躍的に上昇させたが、反面従来の情報検索システムの持つ様々な問題点が露呈されつつある。とりわけ、個人が得られる情報量の増加に対しそれらを処理する有効な手段がないために発生する情報洪水は大きな問題となっており、これを解決するための情報検索方式の確立が急務となっている。

ここ数年のインターネットの大衆化はまた情報検索に関して初心者であるユーザに処理不可能な程の大量の情報を供給するという結果を生み出している。インターネットを用いた情報検索技術は急速に実用化が進みつつあり、今日様々な検索エンジンが稼働中であるものの、その多くは文字列の一致や人手によるカテゴリ分類といった旧来の技術に基づいており、得られる結果はタイトルの羅列であることが多い。これらの検索システムでは、得たい情報を得るための適切なキーワードの選択が難しい、出現する検索結果の情報量を予測することが困難である、得られた結果から必要なものを選択するのが大変である等の問題点を依然抱えている。

これらの問題を解決するアイデアとして、我々は従来の検索システムが生成するタイトルの一次的な羅列からなる検索結果ではなく、各文書がオブジェクトとして表現されている仮想的な三次元空間を用いた検索結果の表示を考案した。このとき文書同士の内容の解析を行ない内容的に近いと思われる文書は仮想空間上でお互いに近くに、遠いと思われる文書は仮想空間上でお互いに遠くに置くような計算方式を採用する。さらにこれらの文書を仮想空間上で検索する際に手がかりとなるように、文書の特徴づける単語を自動的に抽出し、文書の配置を決定すると同様のメカニズムでこれらの単語の仮想空間上での配置を決定しオブジェクトとして表現した。

さらに我々はこのようにして仮想空間上に自動的に配置された文書オブジェクト、単語オブジェクトをユーザが自由に配置を変えられるような機構を追加した。位置を変更されたオブジェクトはそれに関連するオブジェクトに影響を与え内容的に近いオブジェクトは近くに遠いオブジェクトは遠くに配置されるように再計算される。例えばユーザにとって興味のあるオブジェクトを手前に、興味のないオブジェクトを奥に置くことによりユー

ザの興味を反映した情報空間を容易に構成することができる。

我々は上述のアイデアに基づくプロトタイプシステム 3D Stroller を開発した。本論文ではその概要を紹介し、このシステムの持つ自己組織化機構について解説し、関連する研究との差異について述べる。

2 システムの概要

3D Stroller の基本的なアイデアは仮想的な三次元空間に文書の集合から得られる情報の構造を提示し、ユーザがこれらの情報空間を直接操作できるようにするということである。文書、単語は仮想空間上でそれぞれオブジェクトとして表現され、仮想空間上での距離が内容的な関連の遠近を反映するようになっている。(図3)

3D Stroller はまたユーザとのインタラクションを通じて得られた統計的な情報を可視化してユーザに提示する。単語、文書オブジェクトはユーザの好みによって配置を変えることができ、それによって情報空間の再構成が行なわれる。これらの結果として得られる構造はユーザの好みを反映したものとなっている。例えば、もしユーザが興味のある手前にオブジェクトを持ってくるとそのオブジェクトに関連の深い他のオブジェクトが近傍に寄って来るとともに、関連の浅いオブジェクトは遠くに追いやられる。この再構成プロセスはユーザにとって必要な情報の発見を容易するとともにユーザにとって不必要な情報のフィルタリングを可能にしている。時にはユーザにとって思いもかけない情報の構造が得られることもある。

3D Stroller におけるインタラクションのプロセスにおいて、文書と単語は三次元仮想空間上のオブジェクトとして表現され、ユーザはその好みに応じてオブジェクトを移動させることができる。すなわち、興味のあるオブジェクトを近く、興味のないオブジェクトを遠くに置くあるいは削除し、ユーザ自身の価値判断で近いオブジェクト同士を近くに寄せ集める等の操作が可能になる。こうしてユーザの思考方法に従ってユーザ自身の個人的な情報空間が構成される。

こうして得られた構造はユーザが興味を持つ文書、単語オブジェクトから検索式を生成することによって情報検索に役立てることができる。生成された検索式はインターネット上の検索エンジンに投入され、これによって得られた文書集合は本システ

ムを持つデータベースに格納される。格納された文書は仮想的な三次元空間に選択的に登場する。これらのプロセスは巨大なデータベースに対し絶えず検索式を自動生成し検索を行なうシステムととらえることが出来る(図1)。

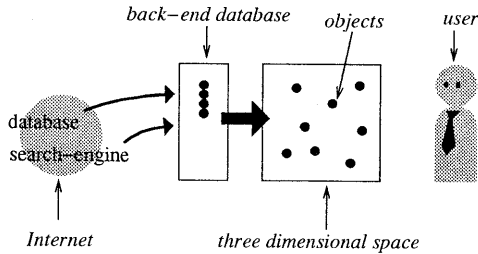


図1: システムアーキテクチャ

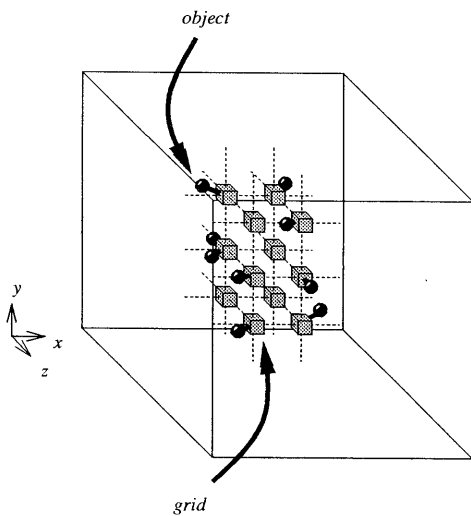


図2: 三次元空間の構造

3 自己組織化のメカニズム

3D Strollerの提供する三次元仮想空間では文書と単語はいずれもユーザが操作可能なオブジェクトとして表現される。文書オブジェクトはその文書を含む単語とその頻度のペアの集合から成るベクトルを持っている。単語はそれ自身と頻度1のペアを持つ特殊なベクトルを持つ。一方三次元

空間内の各格子点もこれらと同様に単語とその強さのペアの集合から成るベクトルを持っている(図2)。

各文書オブジェクト、各単語オブジェクトは最も類似する格子点に配置される。このときの類似度の計算はグリッドとオブジェクトが持つベクトルの内積を取ることで行なう。格子点のベクトルは配置された文書オブジェクトの持つベクトルに接近するように変更されるが、このときに近傍のグリッドのベクトルも若干接近するように変更される。これらのプロセスにより配置された文書オブジェクトに類似した文書オブジェクトが近傍に現れやすくなる。ユーザがオブジェクトの位置を変更するとその近傍のグリッドがオブジェクトの影響を受ける結果似ているオブジェクトが接近し似ていないオブジェクトが離れる。これを繰り返すとユーザの好みに応じた情報空間が生成される。

どのグリッドとも閾値以上に近くない単語オブジェクト、文書オブジェクトは三次元空間から自動的に削除される。一方、三次元空間上に存在するオブジェクトをもとに検索式が自動生成され、これに比較的近い文書がデータベースから検索され三次元空間上に登録される。

上述のように絶えず動的にオブジェクトを入れ換えつつ配置の計算を行なうという自己組織化のための処理負荷は非常に重く、通常の実装方式ではターンアラウンド時間が長くシステムのインタラクティブ性を保てない。そこで我々はどのオブジェクトを先に処理するかを決定する際にユーザの興味を考慮に入れた。すなわち、ユーザが動かしたオブジェクトと移動先のグリッドを優先して処理することとした。また、これらの近傍のグリッド、オブジェクトも順次優先度を若干上げて処理することとした。これらの実装によりインタラクティブ性を保ちつつ自己組織化の処理を進めるシステムの構築に成功した。

このシステムの実装方式にはユーザの最後の入力からの経過時間を考慮に入れている。経過時間が短ければユーザがオブジェクトを動かしている最中であるとし、経過時間が長ければオブジェクトを観察していると想定した。前者であればユーザとのインタラクションの処理を優先させ、後者であれば自己組織化の処理を優先させることで、インタラクティブ性と自己組織化の処理を両立している。

4 関連研究

Galaxy of News system [4] は大量の文書を可視化するための三次元的なアプローチである。このシステムの提供する情報空間は、キーワードの階層関係、記事の見出し、記事の本文の三つの層から成っている。X 軸、Y 軸はキーワードを区別して置くために用いられ、Z 軸はキーワード、見出し、本文という情報の詳細さを示すために用いられる。3D Stroller は Z 軸を主にユーザの興味を示すために用いており、ユーザは情報空間内で文書オブジェクト、単語オブジェクトを自由に置くことが可能となっている。情報空間をインタラクティブに変更可能な点が 3D Stroller の特徴の一つである。

Scatter/Gather [1, 2, 3] は大量の文書を対象にしたクラスタベースのブラウジング技術である。このような技術ではクラスタ間の関係をどのように表現するかがキーとなる。3D Stroller はグリッドを用いて多数のクラスタを生成し、これらを情報空間の鳥瞰する手段として可視化することができる。3D stroller は内容的に近い文書を自己組織化マップ [5] と似た手法で三次元空間内で近い位置に配置する。大量の文書のブラウジングのためには Scatter/Gather のアプローチではクラスタの階層を作らざるを得ず、処理負荷の増大によりインタラクティブ性を保てないことが予想される。3D stroller はターンアラウンド時間を短縮するため優先順位に基づく処理のスケジューリングを行っており、ユーザが興味を持っていると予想されるオブジェクトを優先して計算することでインタラクティブ性を確保している。

自己組織化マップ [5] は入力ベクトルの集合をその相互の距離を出来るだけ保ったまま二次元のグリッドに割り当てるニューラルネットワーク的な手法である。[6] は自己組織化マップを情報検索に適用した例である。3D stroller はオブジェクトの初期配置に関してこれらに近い手法を用いている。しかしながらこれらの手法は基本的には静的な入力ベクトルの集合を想定しており動的に入力ベクトルが変更されたりオブジェクトの位置が変更されたりといった点については考慮されていない。3D stroller はこれらの点を拡張し動的な文書の入れ替えとオブジェクトの再配置を可能にしている。

5 おわりに

新たな情報検索システムのプロトタイプとして 3D Stroller を紹介した。本システムは文書オブジェクト、単語オブジェクトが配置された情報空間を自己組織的に構成するとともにオブジェクトの動的な入れ替えとユーザによる再配置をサポートしておりユーザの個人的な情報空間の構成を容易にしている。本システムの定量的な評価については今後の課題である。

参考文献

- [1] D.R.Cutting et al:
"Scatter/Gather: A Cluster-based Approach to Browsing Large Document Collections", ACM SIG-IR'92, pp318-329
- [2] D.R.Cutting et al:
"Constant Interaction-Time Scatter/Gather Browsing of Very Large Document Collections", ACM SIG-IR'93, pp126-134
- [3] P.Pirolli et al:
"Scatter/Gather Browsing Communicates the Topic Structure of a Very Large Text Collection", CHI'96, pp213-220
- [4] E.Rennison:
"Galaxy of News: An Approach to Visualizing and Understanding Expansive News Landscape2, UIST'94, pp3-14
- [5] T.Kohonen:
"The Self-Organizing Map", Proceedings of the IEEE, Vol.78, No.9, pp1464-1480
- [6] X. Lin et al:
"A Self-organizing Semantic Map for Information Retrieval", ACM SIG-IR'91, pp262-269

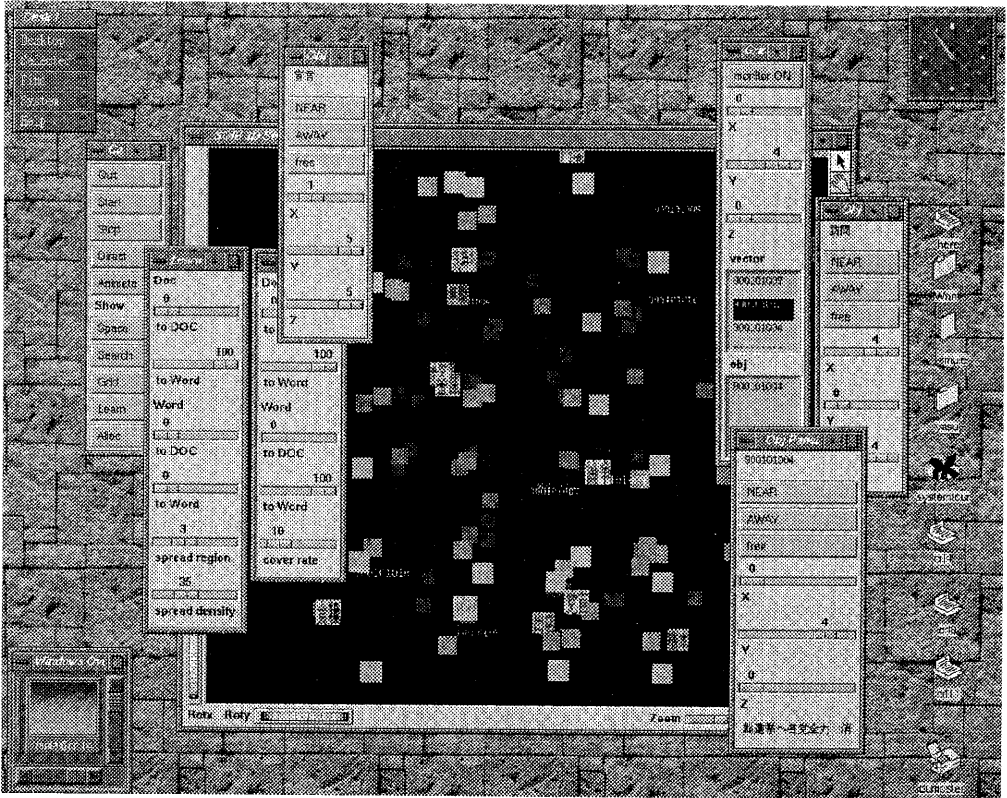


図 3: 3D Stroller プロトタイプシステム