

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2010-15558  
(P2010-15558A)

(43) 公開日 平成22年1月21日(2010.1.21)

(51) Int.Cl. F I テーマコード (参考)  
G06F 3/048 (2006.01) G06F 3/048 651A 5E501

審査請求 有 請求項の数 15 O L 外国語出願 (全 75 頁)

(21) 出願番号	特願2009-147771 (P2009-147771)	(71) 出願人	392026693
(22) 出願日	平成21年6月22日 (2009. 6. 22)		株式会社エヌ・ティ・ティ・ドコモ
(31) 優先権主張番号	08158690.1		東京都千代田区永田町二丁目11番1号
(32) 優先日	平成20年6月20日 (2008. 6. 20)	(74) 代理人	100099623
(33) 優先権主張国	欧州特許庁 (EP)		弁理士 奥山 尚一
		(74) 代理人	100096769
			弁理士 有原 幸一
		(74) 代理人	100107319
			弁理士 松島 鉄男
		(74) 代理人	100114591
			弁理士 河村 英文
		(74) 代理人	100118407
			弁理士 吉田 尚美
		(74) 代理人	100125380
			弁理士 中村 綾子

最終頁に続く

(54) 【発明の名称】 タグクラウドを視覚化する方法、装置およびコンピュータプログラム

(57) 【要約】 (修正有)

【課題】複数の個別タグを備えるタグクラウドを視覚化する方法及び装置を提供する。

【解決手段】タグを複数のスライスに分割するステップと、タグの各スライスがメタ情報測定値と対応するように、タグのメタ情報測定値ベクトルをマッピングするステップと、対応する1つまたは複数のメタ情報測定値に応じて、スライスのそれぞれについて、フォントサイズ、彩度、スライス間の内部スペース、個々のスライスが仮想階段状に置かれたときの実際の高さ、のうちの1つまたは複数を変更することによってタグのメタ情報測定値の時間的変化を視覚化するステップとを備える複数の個別タグを備えるタグクラウドを視覚化する方法。

【選択図】図2

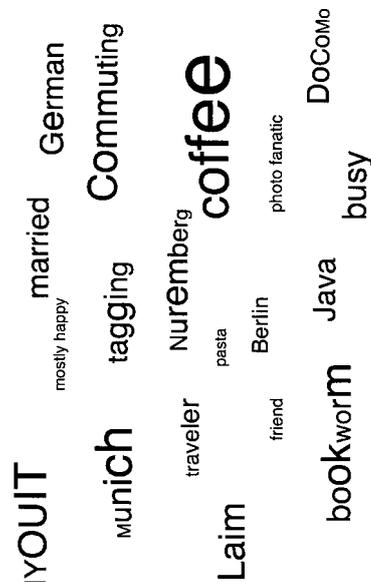


Fig. 2

**【特許請求の範囲】****【請求項 1】**

複数の個別タグを備えるタグクラウド(tag cloud)を視覚化する方法であって、各タグが特定の情報片であり、1つまたは複数の文字で形成されており、タグのメタ情報測定値ベクトルが存在するように一定時間ごとに前記タグのメタ情報が測定され、前記ベクトルの各成分が特定のメタ情報測定値を表し、前記方法が、

タグを複数のスライスに分割するステップと、

前記タグの各スライスが1つまたは複数の前記メタ情報測定値と対応するように、ある時間にわたる前記タグの前記メタ情報測定値ベクトルをマッピングするステップと、

前記対応する1つまたは複数のメタ情報測定値に応じて、前記スライスのそれぞれについて、以下のパラメータ

フォントサイズ、

彩度、

前記スライス間の内部スペース、

1つの固有な重要度測定値が基線に対する実際の高さによって示される、個々の前記スライスが仮想階段状に置かれたときの実際の高さ、

のうちの1つまたは複数を変更することによって前記タグの前記メタ情報測定値の時間的変化を視覚化するステップと、

を備える方法。

**【請求項 2】**

前記メタ情報が、

タグの重要度、

タグ間の相関重要度、

のうちの一方である、請求項 1 に記載の方法。

**【請求項 3】**

あるタグにマッピングした合計期間が前記タグの全長によって図表的に表現される、請求項 1 または 2 に記載の方法。

**【請求項 4】**

予め定められた時間的パターンのどの時間的パターンに各タグが属するかを識別するステップと、

同じパターンに属するタグ同士をグループ分けするステップと、

を更に備える、請求項 1 ~ 3 のいずれか 1 項に記載の方法。

**【請求項 5】**

同一タググループに属する複数のタグが、類似度判定基準によって選択されるか、または、ユーザによって任意に選択され、

前記グループの前記タグについて、メタ情報が、フォントサイズの変化または彩度の変化、のうちの1つ以上によって表示される、

請求項 1 ~ 4 のいずれか 1 項に記載の方法。

**【請求項 6】**

グループを形成する複数のタグが一行に配列され、前記グループがスライスに分割され、各スライスが特定の期間を表し、視覚化されるべきメタ情報のベクトルが、ある時間にわたる前記グループの前記要素間の相関重要度を表しており、前記タググループの視覚表現がある時間にわたる前記グループの前記タグ間の相関関係の変化をフォントサイズまたは彩度の変化に基づいて示すように、前記メタ情報ベクトルが前記タググループの前記スライスにマッピングされる、請求項 1 ~ 5 のいずれか 1 項に記載の方法。

**【請求項 7】**

類似度判定基準に基づいてタググループを選択するステップと、前記タグを一行に配列するステップとを備え、前記グループがスライスに分割され、前記グループの全体形状は、前記類似度判定基準に対応する前記メタ情報の前記時間的変化に基づいている、ステップ、および/または、

10

20

30

40

50

前記グループの前記タグについて、最高個別重要度を有する前記グループのタグが最高フォントサイズまたは最高彩度の位置に配置され、他のタグが前記最高重要度を有するタグに対する相関関係に従って順序付けられるように、配列されるステップ

とを備える、請求項 1 ~ 6 のいずれか 1 項に記載の方法。

【請求項 8】

前記グループのメンバーが一定の類似度を有する基準とされる類似度判定基準に基づいてタググループが選択され、

ある時間にわたる前記グループメンバー間の前記相関関係の全体的相関重要度が、ある時間にわたる前記相関重要度の変化に従って変わるフォントサイズの変化、または、ある時間にわたる相関重要度に従って変わる彩度の変化、によって視覚化される、請求項 1 ~ 7 のいずれか 1 項に記載の方法。

10

【請求項 9】

1 つのタグを基準タグとして選択するステップと、  
前記選択された基準タグと前記他のタグとの相関関係の度合いが、  
前記他のタグのフォントサイズ、または、  
前記他のタグの彩度、

によって表現されるように前記他のタグを表示することによって、前記選択された基準タグと前記他のタグとの相関関係を図表的に視覚化するステップと

を備える、請求項 1 ~ 8 のいずれか 1 項に記載の方法。

【請求項 10】

ある期間にわたるタグの内部重要度がそのフォントサイズおよび彩度のうちの一方によって表現され、

ある期間にわたる前記選択されたタグに対する、あるタグの相関重要度が、前記フォントサイズおよび前記彩度のうちの他方によって表現される、

請求項 9 に記載の方法。

20

【請求項 11】

基準タグが選択され、他のタグは、ある期間にわたる前記選択タグとの個々の相関関係が示されるように同時に表示され、前記相関関係が、

前記相関関係の変化に従って変化するタグのスライス間または文字間のスペース、または

30

前記相関関係の変化に従って変化する彩度、

によって示される、請求項 1 ~ 10 のいずれか 1 項に記載の方法。

【請求項 12】

ある期間にわたるタグと前記基準タグとの前記相関関係を視覚化するステップに加え、前記タグと前記基準タグとの全体的相互関係をさらに視覚化するステップを備える、請求項 11 に記載の方法。

【請求項 13】

前記タグクラウドの前記視覚化により、前記タグクラウド内に表現されたメタ情報に基づいて、前記タグクラウドによって表される内容情報をユーザが照会または閲覧することが可能となり、前記メタ情報は、

40

ある時間にわたるタグの重要度の時間的变化、

ある時間にわたるタグ間の相関重要度の時間的变化、

前記タグクラウドによって図表的に視覚化された前記タグクラウドの前記タグに関するメタ情報、

のうちの 1 つ以上である、請求項 1 ~ 12 のいずれか 1 項に記載の方法。

【請求項 14】

個別のタグを備えるタグクラウドを視覚化する計算装置であって、各タグは特定の情報片であり、1 つまたは複数の文字で形成されており、タグのメタ情報測定値ベクトルが存在するように一定時間ごとに前記タグのメタ情報が測定され、このベクトルの各成分が特定のメタ情報測定値を表し、前記装置が、

50

タグを複数のスライスに分割する手段と、  
 前記タグの各スライスが1つまたは複数の前記メタ情報測定値と対応するように、ある  
 時間にわたる前記タグの前記メタ情報測定値ベクトルをマッピングする手段と、  
 前記スライスごとに1つまたは複数の視覚化パラメータを変化させることによって、前  
 記タグの前記メタ情報測定値の時間的変化を視覚化する手段と、  
 を備え、前記視覚化パラメータは、好ましくは、対応する1つまたは複数のメタ情報測定  
 値に応じて、

フォントサイズ、  
 彩度、

前記スライス間の内部スペース、

1つの固有の重要度測定値が基線に対する実際の高さによって示される、個々の前記  
 スライスが仮想階段状に置かれたときの実測の高さ、  
 のうちの1つまたは複数を用意している、装置。

10

#### 【請求項15】

コンピュータプログラムコードを含むコンピュータプログラムであって、前記コンピ  
 ュータプログラムコードがコンピュータ上で実行されると、請求項1～13のうちのいずれ  
 か1項に記載の方法を前記コンピュータが実行することができる、コンピュータプログラ  
 ム。

#### 【発明の詳細な説明】

#### 【技術分野】

20

#### 【0001】

本発明はタグクラウド(tag cloud)を視覚化する方法および装置に関する。

#### 【背景技術】

#### 【0002】

タグクラウドは、集められた複数のタグ内の各タグに(所与のデータセット内での頻度  
 に基づいて)重要度値を割り当てることによって(構造化されていない)データを表現す  
 る一般的な方法になってきている。いわゆるタグクラウドとは、特定の順序の有無にか  
 わらず、多数のタグの集合体と見なすことができる。例えば、クラウド内のタグはアル  
 ファベット順に並べられていることが多い。タグの重要度は各々のタグに用いられるフォ  
 ントサイズによって示されるので、あるタグについて、重要度が低いほど小さいフォ  
 ントサイズで示される。なお、タグは、特定の情報片を参照するキーワードであるから、シ  
 ンプルな分類や類別が可能となる。以下の記載中、「タグ」という用語は、クラウド構成内  
 に示すことができ且つ基本的にまず見ることができない数個の属性によって限定すること  
 もできる、いずれか1個の情報に対する包括的用語として用いられる。今日、タグクラ  
 ウドの主用途は、後から参照できるように特定のキーワードで予めマーキング(タグ付  
 け)された人気の高いデータ項目を探すこと、および、大きなデータセット内の人気の高  
 いデータ項目をそのデータセット内での頻度に基づいて表示すること、である。それ  
 ぞれのタグに用いられるフォントサイズによってタグの重要度が示される従来のタグ  
 クラウドの例が図1に示されている。

30

#### 【0003】

その人気の高さにもかかわらず、タグクラウドは、視覚化できるメタデータ(実際の情  
 報に関する記述)の量に関してにかなり限定されており、表現されたデータセット内を  
 ナビゲートする限られた手段であるに過ぎない。一般的なタグクラウド視覚化技術の  
 大きな欠点の1つは、特定の瞬間の所与のデータセットのスナップショットしか表示  
 できないため、データセットの最新サンプルしか表現できないことである。その結  
 果、現在のタグクラウド視覚化では、ある時間にわたる1つまたは複数のタグの重  
 要度に関して一定の結論を引き出すことができない。このように、特定の瞬間の  
 人気の高い項目しか強調表示できないので、動的で履歴的な情報表現へのタグ  
 クラウドの応用性は限られている。これまで、ある時間にわたるデータセット内の  
 重要度および相関分布が考慮されたことはなかった。

40

50

## 【0004】

非構造化情報の視覚化および表現に主として用いられるクラウド内の各タグは、同一クラウド内の他のタグとは無関係に処理および表現される。これでは、基本的な共起度(co-occurrence)測定または潜在的に更に高度な意味論的類似性を示すタグ間の単純な相関関係さえも表現することができないため、タグ間の相関関係が各シングルタグの視覚的重要度に影響を及ぼすことはない。同一クラウドのタグ間の一定の相関関係や連結性を表示する可能性が現状のタグクラウドに欠如していることは明らかである。従って、タグクラウドの範囲は、暗黙の情報を明示的に視覚化することに限定され、表現されたデータとインタラクション(interaction)するごく基本的な手段でしかない。

## 【0005】

これまで、ある時間にわたる重要度および相関分布に基づいてデータセットの履歴サンプルをナビゲートしたり照会(query)を制限したりするために用いることもできる時間的コヒーレンスを識別することを目的として2次元(2D)タグクラウド内の複数のシングルタグまたは複数のタググループ内の時間的重要度および相関分布を視覚化できるソリューションは、知られていない。基本的に既存のクラウド表現は、ある瞬間の所与のデータセット内のシングルタグの頻度を視覚化することに限定されている。その結果、所与のデータセット内をナビゲートするためのタグクラウド適用範囲は非常に限られている。

## 【0006】

非特許文献1において、著者らは、「類似している」と識別されてきたクラウド構成内のタグを相対的な共起度測定値に基づいてグループ分けするクラスタリング技術を提案している。この手法は類似タグを視覚的に識別するのに役立つが、タグの重要度分布ならびにタグの相関分布に関する時間的側面は考慮されていない。また、提案された視覚化技法は、所与のデータセットのナビゲーションに関するタグクラウドの機能性を向上させるものではない。また、グループ分けアルゴリズムでは、一般的な動向(trend)、すなわちある時間にわたる相関分布ではなく、共起度測定値しか考慮されない。

## 【0007】

非特許文献2で著者らは、クラウド内の白色スペースの量と配置を最適化することについて、既存のタグクラウド視覚化技術を向上させるためのアルゴリズムに注目している。著者らは、シングルタグの視覚的な重みおよび重要度を示すためにフォント間隔(spacing)属性を利用することも提案しているが、タグ内のスペースは一定のままであり、時間的重要度または相関分布を符号化するものではない。共起度測定に基づいたタグのグループ分けがクラウド内に適用される。

## 【0008】

非特許文献3には、ある時間にわたってタグを視覚化しようという試みが記載されている。著者らは動画を用いた方法を採用しているが、これが彼らの方法を静的な2次元クラウドに適用できない理由となっている。基本概念は、タイムスライダによって選択された既定の時間間隔を表現するようになっているタグ選集を提示することである。しかし、この手法は、ある時間にわたってタグを記述することではなく、(タグ選集によって)一定期間を特徴づけることに関するものである。また、タグとのインタラクションによる照会制限を実現できない。このタグ選集は、所与の時間間隔内のタグ頻度にもみ基づいている。

## 【0009】

非特許文献4で、著者は基本的な共起度測定法を超える類似性測定法を定義している。この類似性測定では、実際のタグコントリビュータ(contributor)およびこのコントリビュータの過去のタグ選集に関する情報が考察されている。その結果、距離に基づいた類似タグの視覚化が提案されている。繰り返すが、実際のタグ視覚化についても、提案されるクラウドとのインタラクションパターンについても、ある時間にわたるタグの重要度および相関分布に関する時間的側面は考慮されない。従って、タグの時間的側面(タグの動向(trend)、すなわちある時間にわたる相関分布)に基づくタグのグループ分けは適用できない。

10

20

30

40

50

【先行技術文献】

【非特許文献】

【0010】

【非特許文献1】Yusef Hassan-Montero and Victor Herrero-Solana, "Improving Tag Clouds as Visual Information Retrieval Interfaces", International Conference on Multidisciplinary Information Sciences & Technologies, 2006

【非特許文献2】Owen Kaser and Daniel Lemire, "Tag-Cloud Drawing: Algorithms for Cloud Visualization, Tagging and Metadata for Social Information Organization", World Wide Web Conference, 2007

【非特許文献3】Micah Dubinko, Ravi Kumar and Joseph Magnani, "Visualizing Tags over Time", World Wide Web Conference, 2007 10

【非特許文献4】Blake Shaw, "Utilizing Folksonomy: Similarity Metadata from the Delicious System", <http://www.metablake.com/webfolk/web-project.pdf>

【発明の概要】

【発明が解決しようとする課題】

【0011】

本発明が解決しようとする課題は、複数の個別タグを備えるタグクラウドを視覚化する方法および装置を提供することである。

【課題を解決するための手段】

【0012】

一実施形態によれば、複数の個別タグを備えるタグクラウドを視覚化する方法が提供される。ここで、各タグは特定の情報片であり、1つまたは複数の文字で形成されており、タグのメタ情報測定値ベクトルが存在するように一定時間ごとに前記タグのメタ情報の測定が行われており、前記ベクトルの各成分はある時刻のあるメタ情報測定値を表す。この方法は、

タグを複数のスライスに分割するステップと、  
前記タグの各スライスが1つまたは複数の前記メタ情報測定値と対応するように、ある時間にわたる前記タグの前記メタ情報測定値ベクトルをマッピングするステップと、

前記対応する1つまたは複数のメタ情報測定値に応じて、前記スライスのそれぞれについて、下記パラメータ

フォントサイズ、  
彩度、

前記スライス間の内部スペース

のうちの1つを変化させることによって、前記タグの前記メタ情報測定値の時間的変化を視覚化するステップと

を備える。

【0013】

このようにして、特定のタグについてある時間にわたって記録されたメタ情報をタグクラウド内に視覚化することができるので、結果として得られたタグクラウドによって、ある時間にわたって測定（または入力）データがどのように変化するかに関する時間的情報が提供される。従って、このような時間的側面に基づいて前記データの照会および閲覧を可能にするインタフェースとしてタグクラウドを用いることができ、それによって照会および閲覧の目的のためにタグクラウドによって表示されるデータとのインタラクション法およびそのようなデータの処理法の新たな次元が開かれる。

【0014】

一実施形態によれば、前記メタ情報は、タグの重要度およびタグ間の相関重要度のうちの一方である。これら2種類の情報は、データセット閲覧時またはデータ照会時のユーザにとって深い関心事であるので、これらのうちの一方または両方が表示できれば有利である。

【0015】

20

30

40

50

一実施形態によれば、あるタグにマッピングした合計期間は前記タグの全長によって図表的に表現される。

【0016】

これにより、タグクラウド内に示されるある時間の時間範囲に関する印象がユーザに与えられる。

【0017】

一実施形態によれば、本方法は、予め定められた複数の時間的パターン(temporal pattern)のうち、各タグがどの時間的パターンに属するかを識別するステップと、同一パターンに属するタグをグループ分けするステップとを含む。

【0018】

これによって、同じ時間的パターンを示すタグをユーザが識別しやすくなり、ユーザはそのうちの1つの選択できる。また、例えば、グループ内のタグについて詳細情報を読み出すことができる特定グループを選択することにより、このようなグループ分けを閲覧または照会に用いることもできる。

【0019】

一実施形態によれば、同一タググループに属する複数のタグは、類似度判定基準によって選択されるか、またはユーザによって任意に選択され、前記グループのタグについて、メタ情報は、

フォントサイズの変化、

彩度の変化、

のうちの1つまたは2つ以上によって表示される。

【0020】

ユーザまたは類似度判定基準(ある時間にわたる重要度の類似形状など)によるタグ選択によるタグのグループ分けは理にかなっており、その後、このグループのタグについて、フォントサイズまたは彩度を変化させることによってメタ情報を表示させてもよい。

【0021】

一実施形態によれば、グループを形成する複数のタグが1列に配列され、前記グループがスライスに分割され、各スライスはある期間を表し、視覚化されるべきメタ情報のベクトルは、ある時間にわたる前記グループの前記要素間の相関重要度を表しており、前記メタ情報のベクトルは、前記タググループの可視表現がフォントサイズまたは彩度の変化に基づいてある時間にわたる前記グループのタグ間の相関関係の変化を示すように、前記タググループの前記スライスにマッピングされる。

【0022】

これにより、ある時間にわたるタググループの要素間の全体的相関関係の時間的変化を視覚的に表示することが可能となる。

【0023】

一実施形態では、これによって、相関重要度が最高の場所(フォントサイズが最大となる位置または彩度が最高となる位置を意味することもある)に、最高個別重要度を有する個別タグが配置されるように前記グループ内のタグの配列順序が選ばれる。

【0024】

一実施形態によれば、本方法は、類似度判定基準に基づいてタググループを選択するステップと、前記タグを1列に配列するステップとを備え、前記グループはスライスに分割され、グループの全体形状は、類似度判定基準に対応するメタ情報の時間的変化に基づいており、および/または、グループのタグは、最高個別重要度を有するグループのタグが最高フォントサイズまたは最高彩度の位置に配置されるとともに、他のタグが最高重要度を有するタグに対する相関関係に従って順序付けられるように、配列される。

【0025】

グループの全体形状が類似度判定基準の時間的変化によって決まることにより、類似度判定基準(例えば、「上昇型」または「下降型」など、個別タグの重要度の特定パターンであってもよい)がどのように見えるか、すぐに分かるようになる。また、最高重要度を

10

20

30

40

50

有するタグとの相関関係に従ってタグを一列に配列することによって、相関関係情報までも視覚化できる。

【0026】

一実施形態によれば、タググループは、前記グループのメンバーが一定の類似度を有する基準とされる類似度判定基準に基づいて選択されている。ある時間にわたるグループメンバー間の相関関係の全体相関重要度は、ある時間にわたる相関重要度の変化に従って変わるフォントサイズの変化、または、ある時間にわたる相関重要度に従って変わる彩度の変化、によって視覚化される。

【0027】

このようにして、グループのメンバー間の全体的相関関係を視覚化できる。

10

【0028】

一実施形態によれば、本方法は、1つのタグを基準タグとして選択するステップと、前記選択された基準タグと他のタグとの相関関係の度合いが、前記他のタグのフォントサイズ、または、前記他のタグの彩度、によって表現されるように他のタグを表示することによって、前記選択された基準タグと前記他のタグとの相関関係を図表的に視覚化するステップとを備える。

【0029】

このようにして基準タグと他のタグとの間の相関関係を視覚化できる。

【0030】

一実施形態によれば、ある時間にわたるタグの内部重要度が、タグのフォントサイズまたは彩度のうち的一方によって表現され、ある時間にわたる前記選択されたタグに対するタグの相関重要度が、前記フォントサイズまたは前記彩度のうち他方によって表現される。このようにして、タグ自体の重要度および基準タグに対するこのタグの相関重要度の両方ある時間にわたって視覚化できる。

20

【0031】

一実施形態によれば、本方法は、選択された基準タグと他のタグとが、ある時間にわたる前記選択されたタグに対する個々の相関関係が示されるように、同時に表示され、前記相関関係が、前記相関関係の変化に従って変化するタグのスライス間または文字間のスペース、または、前記相関関係の変化に従って変化する彩度、によって示される。

【0032】

30

一実施形態によれば、本方法は、タグと前記基準タグとのある時間にわたる相関関係を視覚化するステップに加え、前記タグと前記基準タグとの全体的相関関係をさらに視覚化するステップを備える。

【0033】

これにより、ある時間にわたって変化するあるタグと基準タグとの相関関係に加えて、あるタグと基準タグとの全体的相関関係を視覚化することも可能となる。全体的相関関係は、例えば、タグの近くに配置された単線の伸び具合によって表示してもよい。

【0034】

一実施形態によれば、タグクラウドの視覚化により、前記タグクラウド内に表現されたメタ情報に基づいて、前記タグクラウドによって表される内容情報をユーザが照会または閲覧することが可能となり、前記メタ情報は、

40

ある時間にわたるタグの重要度の時間的变化、

ある時間にわたるタグ間の相関重要度の時間的变化、

前記タグクラウドによって図表的に視覚化された前記タグクラウドのタグに関するメタ情報、

のうちの1つ以上である。

【0035】

一実施形態によれば、複数の個別タグを備えるタグクラウドを視覚化する装置が提供される。ここで、各タグは特定の情報片であり、1つまたは複数の文字で形成されており、タグについて、前記タグのメタ情報測定値ベクトルが存在するように一定時間ごとにメタ

50

情報の測定が行われており、前記ベクトルの各成分は特定のメタ情報測定値を表す。前記装置は、タグを複数のスライスに分割する手段と、前記タグの各スライスが1つまたは複数の前記メタ情報測定値と対応するように、ある時間にわたる前記タグの前記メタ情報測定値ベクトルをマッピングする手段と、前記対応する1つまたは複数のメタ情報測定値に応じて、前記スライスのそれぞれについて、フォントサイズおよび彩度というパラメータのうちの1つを変化させることによって、前記タグの前記メタ情報測定値の時間的変化を視覚化する手段とを備える。

【0036】

このようにして、本発明の一実施形態に従う装置を実施できる。

【0037】

別の実施形態によれば、本装置は、本発明の実施形態のうちのいずれか1つに従う方法を実施する手段を備える。

【0038】

本発明のさらに別の実施形態によれば、実行時に本発明のいずれかの実施形態に従う方法をコンピュータが実行することができるコンピュータプログラムコードを含むコンピュータプログラムが提供される。

【図面の簡単な説明】

【0039】

【図1】業界公知のタグクラウドを示す図である。

【図2】本発明の実施形態に従うタグクラウド表現の略図である。

【図3】本発明の別の実施形態に従うタグクラウド表現の略図である。

【図4】本発明のさらに別の実施形態に従うタグクラウド表現の略図である。

【図5】本発明のさらに別の実施形態に従うタグクラウド表現の略図である。

【図6】本発明のさらに別の実施形態に従うタグクラウド表現の略図である。

【図7】本発明のさらに別の実施形態に従うタグクラウド表現の略図である。

【図8】本発明のさらに別の実施形態に従うタグクラウド表現の略図である。

【図9】本発明のさらに別の実施形態に従うタグクラウド表現の略図である。

【図10】本発明のさらに別の実施形態に従うタグクラウド表現の略図である。

【図11】本発明のさらに別の実施形態に従うタグクラウド表現の略図である。

【図12】本発明のさらに別の実施形態に従うタグクラウド表現の略図である。

【発明を実施するための形態】

【0040】

本発明の実施形態に従う概念は、相関関係という側面のみならず時間的な側面を考慮することによって、データの視覚化およびナビゲーションに対するタグクラウドの基本概念を大きく広げるものである。以下の記載で明らかになるであろうが、ここで提案される視覚化およびインタラクション法によれば、大きなデータセットを表現したり照会したりするための新しい方法が実現される。

【0041】

詳細には、ある時間にわたる1つのタグおよび複数の相関タグの重要度を強調表示するためにタグクラウド内の付加的な情報(メタ情報)を符号化する新規な視覚化技術に基づいて、大きなデータセット内の重要データ項目をどのように視覚的に記述してどのように閲覧に役立てることができるかを説明する。

【0042】

メタ情報は、例えばGPSセンサからのデータを一定間隔で記録することによって自動的に記録できる。このデータに基づいて位置情報が判断され、また、この位置情報に基づいて、例えば、その位置の都市名、場所の名称(例えば、事務所)等といった、対応タグを自動的に記録することもできる。なお、タグは、ユーザによって入力されたメタ情報であってもよく、例えば、ユーザのそのときの気分(だるい、幸せ、悲しい、等)やユーザの行動(運動、読書、歩行)など、ユーザの状況を表すメタ情報に基づくものであってもよい。そのようなメタ情報がユーザによって入力され、その後、対応する時間と一緒に記

10

20

30

40

50

録されてもよい。従って「測定」という用語は、ある時刻にユーザがメタ情報を入力することも、ある時点におけるメタ情報の「測定」と見なされる場合があることを理解されたい。

【0043】

ユーザ入力に由来するものであっても、自動測定に由来するものであっても、そのようなメタ情報はすべて、数種類または数タイプのメタ情報からなる集約体となり、異なる種類または異なるタイプは対応するタグによって表現され、ある時間にわたってタグの重要度（例えば、一定時間間隔内の発生回数によって）が記録される。

【0044】

このようにして収集された、ある時間にわたる内容情報を表す内容データ（すなわち「測定データ」）のプールは、ユーザがこの情報プールを有意味に表示、閲覧、および照会できるようにするために本発明の視覚化メカニズムを適用するためのベースとなり得る。内容情報は全く異なっていることがあるし、時間的情報も含んでいるので、そのような情報プールを表示および照会するためには現在のインタフェースは適していない。以下に記載の本発明の各実施形態によれば、これまで不可能であった態様によりユーザがこのデータにアクセスし、照会し、閲覧するのを支援することができる。

10

【0045】

メタ情報の収集に使用できるツールは、携帯電話、スマートフォン等の携帯デバイスとすることができる。このような携帯デバイスは、自動測定（位置測定、またはデバイスのセンサに由来する他のデータ測定）を実行できるようになっていてもよいし、ユーザがメタ情報を入力し、入力されたメタ情報が対応する時刻と一緒に記録されるようになっていてもよい。その後、携帯デバイスは、対応する機能を実行するコンピュータプログラムを実行することによって本発明の実施形態に記載の視覚化方法を実行してもよいし、対応タグに対応し、対応タグによって表現されるメタ情報のある時間にわたって記録する「測定」によるデータ収集によって得られたメタ情報をユーザが照会および閲覧できるようにしてもよい。

20

【0046】

一実施形態において提案される手法によれば、1つのタグの表現内でそのフォント特性を変化させる（例えば、種々のフォントサイズを使用することなどによって）ことで、ある時間にわたるタグの重要度分布を記述できる。従って、タグの重要度が著しく減少した場合、固有(distinct)タグのフォントサイズも小さくなるし、重要度が増加した場合は固有タグのフォントサイズも大きくなる。その結果、観測対象の複数タグの重要度の動向を1つの（静的な）クラウド内に視覚化することができ、従って、現時点の重要度のみならず、ある時間にわたるその重要度分布にも基づき、タグをより確実に識別および選択することが可能となる。また、複数のシングルデータ項目や複数項目から成る複数グループの動向を視覚的に識別できるので、提供された情報を深く理解できるようになることに加え、過去に記録された該当データの履歴サンプルを閲覧することも可能となる。1つのタグ表現の固有部分を選択することによって、このデータ項目の履歴サンプルを明らかにすることができる。同様に、ある時間にわたるタグの重要度分布を視覚化することにより、所与のデータセットを、動向という観点から閲覧することが可能となる。例えば、ある時間にわたって望ましい動向（例えば、重要度の増加）を有するタグを選択することができ、また、ある時間にわたって同じ動向を有する他のタグ（および、その関連情報）をサーチすることができる。従って、本発明の実施形態による視覚化法は、ある時間にわたって記録されたデータを視覚化することのみならず、新規なインタフェースによってそのようなデータを照会および閲覧できるようにするものでもある。

30

40

【0047】

今日、クラウド内のタグの重要度は、所与のデータセット内の頻度に基づいたものに過ぎない。しかし、時として、タグの重要度および重要度の適切な指標は頻度だけではない。そうではなく、タグの重要度は、同一クラウド内の他のタグとの相関関係すなわち結びつき、ならびに、ある時間にわたる相関関係の重要度によっても、大きく左右される。2

50

つ以上のタグの相関度が共起度測定値に基づくこともあり、タグの意味論的または統計的類似性（近さ）を示す場合もある。既存のタグクラウド表現は相関関係の度合いを視覚化することができないため、1つのクラウド内のタグ同士を区別するための適切な区別因子が欠如している。その結果、データセットをナビゲートすることもできないし、相関性のある数行のタグに沿った所与の照会を制限することもできない。

#### 【0048】

本発明の実施形態による視覚化法およびインタラクション法は、固有のグループ分け、色分け、または距離表示により、タグ間の相関関係因子を視覚化することによって、この制約に取り組むものである。ある時間にわたるタググループの相関重要度は、その関連重要度と一致するようにグループ全体の形状を調整することによって視覚化できる。色分けパラダイムにより、彩度100%の基準色で強調表示された特定のタグを選択することが可能となる。一実施形態において、すべての相関タグが同一色調で強調されるが、その相関度に基づいて、相関度が高いほど彩度が高くなるように彩度選択が行われる。一実施形態において、タグ内の白色スペースを調節することによって、ある時間にわたる相関度を表現できる。また、さらに別の実施形態では、所与の2つのタグの全体的相関度について、最新の相関度に高い強調が施されてもよい。

10

#### 【0049】

提案される視覚化法およびインタラクション法では、高度なタグクラウド表現を生成することにより、大きなデータセットをナビゲートおよび閲覧する新たな方法が可能となる。照会は、複数のシングル項目または複数のタググループの重要度分布に基づいて制限可能であり、関連項目による閲覧を可能にし、所与のデータセットの履歴サンプルをナビゲートする手段を提供する。従って、ある時間にわたって識別された、時間的な条件、特定の相関関係従属性、または類似した動向に基づいて、照会を制限することもできる。

20

#### 【0050】

提案される視覚化およびインタラクションパラダイムでは、データセット内の重要項目は、その時点の重要度だけでなく、ある時間にわたるその重要度およびある時間にわたる他のタグとの相関関係の重要度に基づいて、強調表示することが可能である。従って最近のみ人気の高い（短期間）タグを切り離すことができ、より長期間にわたって高い相関度を備えた重要なタグが高い視覚的に重みを得る。また、提案される手法によれば、ある時間にわたるその重要度および相関関係の分布に基づいて、複数のシングルデータ項目または複数の項目グループを効率的に選択することが可能となり、また、過去に記録された履歴データサンプルを照会することが可能となり、類似した重要度および相関関係を備えたタグのグループを隔離する手段およびタグ間の相関度に基づいて所与のデータセットを閲覧する手段が提供される。

30

#### 【0051】

ここで、本発明の実施形態を幾分か詳細に説明する。

#### 【0052】

まず、タグ内の重要度分布が視覚化された実施形態を説明する。複数のシングルタグの重要度分布は、一定期間内に記録された重要度測定回数によって定められる。これにより、タグの重要度すなわち関心度の何らかの傾向をある時間にわたって示すことができる。

40

#### 【0053】

提案される視覚化およびインタラクション技法は、既存のタグクラウド表現を著しく拡大し、付加的なメタ情報を静的なクラウド内に視覚化することを可能にし、新規なインタラクションパターンおよび新規な照会制限パターンを可能にするものである。メタデータにより、クラウド内の複数のシングルタグのある時間にわたる重要度分布および複数タグ間のある時間にわたる相関関係を識別および強調表示することが可能となる。

#### 【0054】

提案される視覚化およびインタラクションパラダイムによれば、タグの重要度は、ある瞬間の所与のデータセット内の頻度（1つの単一値）だけに基づくのではなく、ある時間

50

にわたって記録された複数の重要度レベルだけで構成することもできる。このようにして得られる重要度分布により、タグの人気が増加しているか、増加しているか、無変化であるかとは別に、一定期間中のタグの重要度に関する貴重な情報が保持される。提案される視覚化パラダイムによれば、この傾向は、図 2 に示されるような 1 つの静的なタグクラウド内に可視化できる。下記の例では、タグ "coffee" の重要度が増加したことが、大きくなっていくフォントサイズによって示されているが、一方、タグ "commuting" はフォントサイズが小さくなりつつあり、ある時間にわたるその対応重要度が低下している。

【 0 0 5 5 】

所与のタグに使用される実際のフォントサイズおよびその重要度分布は、所与のタグ内の固有セクションの個数（セクションとは、例えば、文字列の一部、すなわちタグの分割可能部分など、タグの文字またはタグの小部分であってもよい）が、対象の重要度測定回数をも制限すると仮定することによって算出できる。

【 0 0 5 6 】

ここで、

【数 1】

$\vec{L}$  : 所与のタグの固有セクションのベクトル

$\vec{S}$  : 所与のタグのある時間にわたる重要度測定値ベクトル

$S_{max}$  : 最大重要度値

$F_{max}$  : 最大フォントサイズ

$F_{min}$  : 最小フォントサイズ

とする。

【 0 0 5 7 】

次式により、任意の重要度値範囲が固有タグの個々のフォントサイズにマッピングされる。フォントサイズは、結果として生じるフォントサイズベクトル

【数 2】

$\vec{F}$

によって示される。

【数 3】

$\vec{F}(\vec{L}, \vec{S}) = [f_0, \dots, f_k]$  ここで

$$\forall_{i=0}^k f_i = F_{min} + \left( \frac{S_i}{S_{max}} \right) \times (F_{max} - F_{min}),$$

$$(F_{max} - F_{min}) > 0,$$

$$\vec{S} = [s_0, \dots, s_k] \text{ 但し } \forall_{j=0}^k s_j, 0 \leq s_j \leq S_{max},$$

$$\vec{L} = [l_0, \dots, l_k]$$

【 0 0 5 8 】

タグの重要度分布を視覚表示することにより、時間的照会制限(temporal query restriction)を定めることができる。タグの水平表示が、観測された時間間隔を表すのに対し、タグの固有セクションは、この時間間隔の時間サブセットを表す。識別可能な時間スライス (= タグのセクション) の個数ならびに実際の重要度測定回数は、特定のタグの全体サイズに依存する場合もある。特定の時間スライスの選択は時間的照会制限として見なすこ

10

20

30

40

50

ともでき、その結果のセットは選択され時間間隔に対応している。このように、観測する時間間隔を狭めることにより、大きなデータセット内の1個の情報を検索することができる。タグのスライスがタグの一文字に対応していてもよいし、文字の一部(「スライス」)に対応していてもよい。前者の場合、重要度の指標は、文字ごとに変化するが単一文字内では同じままである。後者の場合、重要度の指標は文字内で変化することさえある。すなわち、個々の文字のフォントサイズがその文字内であっても変化することさえもある。

【0059】

タグのセクションの個数および時間的測定回数が同じでない場合、すなわち、所与のタグの固有セクションのベクトルの成分数と、ある時間にわたる所与のタグの重要度測定値ベクトルの成分数が異なる場合、一方のベクトルの複数の成分が他方のベクトルの一成分にマッピングされるようにマッチングを行ってもよい。これは、例えば、数個の重要度成分を1つのタグセクションにマッピングさせる場合、複数の重要度の平均重要度値を求めることによって行うことができる。

10

【0060】

タグクラウドのタグによって表現される時間間隔は、予め決めておいてもよいし(例えば、1日、1週間、1カ月等)、ユーザが選択してもよい。例えば、ユーザが自分にとって最も関心のある時間間隔を選ぶこともできるし、収集されたメタ情報に基づいて、重要度データの時間的推移を視覚化するタグクラウドが算出されて表示されてもよい。

【0061】

一実施形態によれば、タグクラウドを形成する全部のタグが、必ずしも同一観測時間間隔を表すわけではないので(例えば、視覚化させようとユーザが選択した間隔全体を網羅する十分なメタ情報データが全タグについて存在するとは限らないので)、一実施形態によれば、対象とされる相対時間のための別の指標が用いられる。この実施形態では、タグの(水平方向の)全長は、重要度測定を展開する全期間の長さである。異なるタグによって別々の時間間隔を表すための長さ変更は、例えば、タグ内の白色スペースを変えることによって行うことができる。図3に示されるように、タグ内の白色スペースを変えるとその水平方向の長さが増し、これは比較的長い期間に相当する。図3に示される例では、例えば、"commuting"というタグは"nuremberg"というタグよりも長い期間をカバーしている。

20

【0062】

この付加的なタグ内の白色スペースすなわちフォント間隔は、次式によって概算できる。

30

【数4】

$$\vec{D}(\vec{T}) = [d_0, \dots, d_k] \quad \text{ここで}$$

$$\forall_{i=0}^k d_i = D_{\min} + \left(\frac{t_i}{T_{\max}}\right) \times (D_{\max} - D_{\min}),$$

$$(D_{\max} - D_{\min}) > 0,$$

$$\vec{T} = [t_0, \dots, t_k] \quad \text{但し} \quad \forall_{j=0}^k t_j \cdot 0 \leq t_j \leq T_{\max}$$

40

ここで：

$d_i$  : タグ内の2つのセクション(例えば、文字)間のスペースすなわち距離

$D_{\min}$  : タグ内の2つのセクション(例えば、文字)間の最小(デフォルト)スペース

$D_{\max}$  : タグ内の2つのセクション(例えば、文字)間の最大スペース

$t_j$  : 関心タグの観測時間間隔

$T_{\max}$  : 特定クラウド内のタグの最大観測時間間隔

50

である。

【0063】

タグ内の白色スペースを変化させる代わりに、異なるフォントタイプすなわち変体を用いることもできる。密集したフォント（例えば、Arial Narrow）によって形成されているタグは短い時間間隔を表し、標準または幅広のフォント（例えば、Arial Regular）は長い期間を表す。

【0064】

上述のタグクラウドの視覚化により、ユーザは、そのようなタグクラウドによって表される内容情報を時間的情報に基づいてナビゲートできるようになる。例えば、時間的変化の観点から上昇傾向を示すタグは最近になって重要度が高まってきているものであるので、ユーザはそのようなタグを選択することによって、時間的情報に基づいてデータセットをナビゲートするための内容情報を選ぶことができる。別の可能性は、1つのタグではなく、このタグによって示される全時間間隔のサブセットである特定の時間間隔を表すタグの一部を選択することである。そのような選択は、例えば、周知の方法でマウスまたはカーソルを用いてタグの選択部分を枠で囲むことによって行うことができる。また、そのように選択された時間間隔は、選択されたタグおよび時間に関連する他の情報など、さらに詳細な情報を入手する際のフィルタすなわち選択判定基準として適用される。一実施形態において、時間のサブセットの選択により、そのように選択された時間のサブセットのタグクラウド全体の再計算をもたらし、それによってタグクラウドを効率的に拡張することもできる。

10

20

【0065】

比較的シンプルな場合、前述の時間的選択および照会制限を既に表現しているタグの固有セクション（例えば、1文字）を選択（例えば、このセクション上でクリックすることによって）できる。また、一実施形態において、ワードプロセッシング時に特定のテキストにマーキングするのと同様に、このタグの複数部分を選択することもできる。この選択は、マウスのクリックによって確定できる。

【0066】

結果として得られる照会を観測期間について制限するためにタグの固有の部分を選択するのは対照的に、一実施形態に従うタグ内スペースの選択には逆の効果がある。この場合、一実施形態によれば、対象期間が延長され、その結果、クラウド内の各シングルタグの再計算がもたらされる。

30

【0067】

ここで、傾向に基づいたタグのグループ分けを用いる別の実施形態を説明する。図2に例示されているようにシングルタグのある時間にわたる重要度分布を計算すると、その（意味論的）類似性に基づいて、または単純な共起度測定値に基づいて固有タグをグループ分けすることができるだけでなく、類似の重要度の動向に基づいてグループ分けすることも可能となる。これについて、数タイプの動向を（一例として）下掲のように識別することができる。

【0068】

【表 1】

bookworm	次式のように重要度が上昇している。 $\forall_{i>0}^l s_i < s_{i+1} < s_l$ (s;固有重要度の測定値)
bookworm	次式のように重要度が下降している。 $\forall_{i>0}^l s_i > s_{i+1} > s_l$ (s;固有重要度の測定値)
bookworm	次式のように重要度の上昇が繰り返されている。 $\forall_{i=0}^{k-1} s_i < s_{i+1} \wedge \forall_{j=k+1}^l s_j < s_{j+1} \wedge s_k > s_{k+1}$
bookWorm	次式のように重要度の下降が繰り返されている。 $\forall_{i=0}^{k-1} s_i > s_{i+1} \wedge \forall_{j=k+1}^l s_j > s_{j+1} \wedge s_k < s_{k+1}$
bookworm	次式のように重要度が下降および上昇している。 $\forall_{i=0}^{k-1} s_i > s_{i+1} \wedge \forall_{j=k+1}^l s_j < s_{j+1} \wedge s_k = s_{k+1}$
bookWorm	次式のように重要度が上昇および下降している。 $\forall_{i=0}^{k-1} s_i < s_{i+1} \wedge \forall_{j=k+1}^l s_j > s_{j+1} \wedge s_k = s_{k+1}$

10

20

30

40

50

## 【0069】

類似した重要度分布を有するタググループを識別したことによって再フォーマットされたタグクラウドが図4に示されている。

## 【0070】

類似した重要度の動向に基づいてタグクラウド構成内の特定グループを識別した後、一実施形態によるグループベースの照会制限により、選択された重要度の動向だけについてののみ、所与の結果セットが再考される。

## 【0071】

一実施形態による、結果として得られたタグクラウド構成の配置は、次のパラダイムによって決まる。各タググループは、原則として、独立したクラウドのように見える。最大タググループは、結果として得られた(複合)タグクラウドの中央に配置され、それに続く全グループが、この中央のグループを囲んで配置される。図4に示される例では、例えば、"laim"および"busy"等のタグを含むクラウドが最大クラウドとして識別されているため、このタグが、結果として得られた複合タグクラウドの中央に配置されている。フィールド数が識別された動向(およびそのようなタググループ)の総数よりも大きくなくてはならない場合、その配置はマトリックス(matrix)に基づく。このように、付加的な視覚的重みが最大すなわち最も重要なタググループに置かれ、これにより、シングルタグが更に迅速にナビゲーション、選択、および識別するのがしやすくなる。図5は、図4に示されるタグクラウドのマトリックス構造を示す。

## 【0072】

以下に、色分けによる相関重要度分布を使用する実施形態を説明する。

## 【0073】

タグの時間的重要度分布に加え、このタグと同じクラウド内の別のタグとの、ある時間

にわたる相関関係もまた、価値のあるメタデータである。時間的相関分布は、指定された時間間隔中に記録される特定タググループの相関関係測定回数と定めることができる。そのような相関関係は、共起度測定結果、意味論的従属性、または、時間、物理的空間、若しくはこのグループ内に存在しない他のデータ項目の通常存在について識別されるこれらの概念間の結びつきを基礎とすることができる。

【0074】

ある時間にわたって共通の重要度分布を有する、識別されたタググループを考えることにより、グループの相関重要度を一列状（あるいは円形状）に視覚化することが可能となる。一列状の場合、列中、x軸の左側が最新の測定のタイムスケールであり、右側が最古の測定のタイムスケールである。一列にグループ分けされたこれら多数のタグは、またも2次元タグクラウドを形成している。グループ内の各タグは、自己の重要度分布に関する固有情報を依然として保持しているが、ある時間にわたるグループの相関重要度は、彩度100%が最高相関重要度を表す、漸次増減配色(fading color schema)によって示される。このようなタグクラウドは、図6に示されるように見えることができる。

10

【0075】

図6のグループに属するタグの選択は、類似の時間的重要度分布だけに基づく選択に限られなくてもよく、例えばユーザによって任意に選択されてもよい。その後、ある時間にわたるタグ間の相関関係は、図6に示されるように、すなわち、時間に対応するx軸に沿って変化する彩度によって、表示してもよい。

【0076】

ここではグループのメンバーを選択するための種々の選択判定基準を記載したが、これらの選択判定基準をどのような組合せで組み合わせることもできることは言うまでもないことを記載しておく。例えば、別の実施形態に記載されるように、タグをその意味論的類似性に基づいて選択できる。これとは別に、またはこれに追加して、一実施形態では、ユーザによる任意選択も常時可能である。

20

【0077】

グループの「相関分布」を表す彩度は次式によって記述できる。

【数5】

$$\forall_{i=0}^k x_i = X_{\min} + \left( \frac{c_i}{C_{\max}} \right) \times (X_{\max} - X_{\min}),$$

30

$$(X_{\max} - X_{\min}) > 0,$$

$$\vec{C} = [c_0, \dots, c_k] \quad \text{ここで } \forall_{j=0}^k c_j, 0 \leq c_j \leq C_{\max}$$

ここで、

$x_i$  : タググループ内の個々のセクションの彩度、

$X_{\max}$  : 最大彩度

$X_{\min}$  : 最小彩度、

$c_i$  : 任意の相関重要度サンプル、

$C_{\max}$  : グループの最大相関重要度

である。

40

【0078】

一実施形態によれば、グループ内のタグの順番も特定のパラダイムに従っている。最も重要なシングルタグが、グループの相関重要度ピーク位置に置かれる（図6では、"coffee"が最も重要なシングルタグとして識別されているので、グループの最大相関重要度と評価される、左から2番目の位置に置かれている）。

【0079】

グループの判定基準： 類似した重要度分布（または任意のタグ選択）

50

タグの形状：個々のタグの重要度分布

タグの配置：最も重要なシングルタグの位置とグループ形状の最大相関重要度の位置と一致する

グループの形状： -

色分け：ある時間にわたるグループの相関重要度を色の漸減によって示す

インタラクション：期間および共通のグループ判定基準（類似した重要度分布）に基づく照会制限

#### 【0080】

ここで、この視覚化パターンをどのようにナビゲーション目的に使えるかを示す例を説明する。例えば、所与の照会をさらに精緻なものとし結果セットを小さくするために、グループの固有部分を選択することができる。これにより、データ項目を記録しておかなければならない期間が、表示された相関重要度の選択領域に対応するように制限される。クラウド内では、期間が識別されるだけでなく、特定のグループも選択されているので、結果として生じる照会は、一実施形態では、共通のグループ判定基準（上の例では、ある期間にわたって類似した重要度分布）によって定められる。

10

#### 【0081】

以下に、グループベースの重要度分布を使用する実施形態を説明する。

#### 【0082】

この実施形態では、グループ内のタグ間の実際の相関関係ならびにグループの共通判定基準は、後からの関心事である。そういうものとしてグループに目を向けると、その共通判定基準ならびにグループ内のタグ間の相関関係を図7に示されるように視覚化できる。

20

#### 【0083】

ここでは、グループの共通判定基準（類似の重要度分布）がグループ全体の形状を形成し、一列に再配列されているが、グループの最も重要なタグ（この場合は"coffee"）が識別され、その後、最も目立つタグとして視覚化される（図7ではフォントサイズで目立たせているが、重要度を表現するために選択されるのであれば、彩度で目立たせることも可能である）。グループ形態内の所与のタグの順序は、グループの最重要タグに対する特定タグの相関度に基づいている。相関度は、グループ内の2つの固有タグ間の従属性を示す。上の例では、タグ"munich"は、"traveler"との相関度よりも、"coffee"との相関度の方が高い。

30

#### 【0084】

形状レンダリングは、第1の例で様式化されたシングルタグの重要度分布のように数学的に記述できるが、タググループは連結された1つのシングルタグとして扱われる。

#### 【0085】

グループの判定基準：類似の重要度分布

タグの形状：グループの形状を決定（従って、形成）する

タグの配置：最も重要なシングルタグがグループの最重要位置に置かれる。残りのタグは、この最重要シングルタグに対する相関度に基づいて順序付けられる。

グループの形状：共通の判断基準（ある時間にわたる重要度分布）がグループの形状を定める

40

色分け： -

インタラクション：期間および共通のグループ判断基準（類似の重要度分布）に基づく照会制限

#### 【0086】

ここで、グループ形状を有する相関重要度分布を使用する実施形態を説明する。

#### 【0087】

ある時間にわたる数個のタグの相関関係の特徴描写することによって、グループの相関関係の重要度がそれなりに示される。従って、グループ内のすべてのタグが、上記の例で論じられたように類似の重要度分布に基づいてではなく、特定の相関関係に基づいて（例えば、意味論的類似性に基づいて）選択されている。提案される視覚化パラダイムは、次

50

いで、ある時間にわたるグループ全体の相関重要度を示すが、それぞれの固有タグの重要度分布は、図 8 に示されるように、もはや主目的ではない。

【 0 0 8 8 】

選択され、図 8 に視覚化されている例に示されているのは、類似タグ（すべてのタグが空間的概念を表現しており、従って、この場合の類似度判定基準は「空間的概念を表現するタグ」である）のグループと、ある時間にわたるグループの相関重要度である。なお、上記グループの重要度は最初に増加し、次に相関重要度が着実に減少している。各シングルタグの形状は、グループ形状にそれなりに適合しており、その結果、グループ全体ならびにある時間にわたるその相関重要度が更に目立つように視覚化されている。やはり、グループ内のタグの順序は特定の原理に従っている。すなわち、最も重要なシングルタグが、最高重要度値が存在する位置を取る。その結果、グループの最重要タグとして"munich"がクラウド構成の第 1 の部分内に置かれ、"nuremberg"、"laim"、および"berlin"がそれに続く。

10

【 0 0 8 9 】

グループの相関重要度の分布は、（上に説明されたように）タグ内重要度分布と同様に記述できるが、結果として生じるフォントサイズベクトルは、この重要度測定によって記述されるのではなく、実際には、グループの相関関係測定によって求められる。

【 数 6 】

$$\vec{F}(\vec{L}, \vec{C}) = [f_0, \dots, f_k] \quad \text{ここで}$$

20

$$\forall_{i=0}^k f_i = F_{\min} + \left( \frac{c_i}{C_{\max}} \right) \times (F_{\max} - F_{\min}),$$

$$(F_{\max} - F_{\min}) > 0,$$

$$\vec{C} = [c_0, \dots, c_k] \quad \text{但し} \quad \forall_{j=0}^k c_j, 0 \leq c_j \leq C_{\max},$$

$$\vec{L} = [l_0, \dots, l_k]$$

30

ここで、

【 数 7 】

$\vec{L}$  : クラウドの固有セクションのベクトル

$\vec{C}$  : ある時間にわたる相関関係測定値ベクトル

$C_{\max}$  : 最大相関値

$F_{\max}$  : 最大フォントサイズ

$F_{\min}$  : 最小フォントサイズ

40

である。

【 0 0 9 0 】

グループの判定基準： 意味論的または統計的類似性

タグの形状： グループの形状に適合（従って、形成）する

タグの配置： グループがその最高相関重要度を有する場所に配置される

グループの形状： ある時間にわたるグループの相関重要度を示す

色分け： -

インタラクション： 期間および共通のグループ判定基準（類似のタグカテゴリ）に基づく照会制限

【 0 0 9 1 】

50

ここで、基準タグ選択および相関度視覚化を使用する実施形態を説明する。

【0092】

ある時間にわたるタグの重要度、タググループの構成、および、ある時間にわたるそれらの相関重要度に注目することに加え、タグ間の相関度は、表現されたデータの別の有益メタ情報である。特定の配色を適用することによってそのような相関関係を可視化し、単純な共起性、意味論的類似性、または統計的解釈を示すことができる。タグ間の相関関係を強調表示するために、基準タグが選択される。特定の基準色を彩度100%のそのフォントに割り当てることにより、基準タグの選択を示す。従って、すべての相関タグは選択タグと同色で強調表示されるが、選択タグとの相関度は相関タグに適用される色の彩度と同等となる。これは、高い相関度は高い色飽和百分率で示されるが、低い色飽和百分率は低い関連度に相当することを意味する。図9は、選択タグとして"coffee"を含むサンプルタグクラウドを示す。このタグクラウド中、"commuting"または"busy"などの相関タグは、漸次増減する彩度によって示されている。

10

【0093】

実際のタグ選択には、2通りの効果がある場合がある。一方では、上述の相関度視覚化により、所与のクラウド内の相関従属性に焦点を当てることができる。別の実施可能パターンは、データセット内における相関関係に基づくナビゲーションである。なぞ、特定のタグが選択されるたびに、相関タグだけを検索するように照会を対応させることも可能である。このように、相関情報のサブセットだけが、有意義な情報を見つけ出すための効率的な手段となると見なされる。

20

【0094】

ここで、時間的相関度の視覚化を使用する実施形態を説明する。

【0095】

タグ内の白色間隔(white spacing)またはフォント間隔を利用することによって、(上に例示したように)シングルタグ表現内に相対的な期間を符号化できるだけでなく、(基準タグに対する)ある時間にわたるその相関度を示すことも可能である。なお、タグ内スペースの増加は相関度の減少を示し、スペースの減少は相関度の増加を示す。タグ内スペースが無変化であることは、ある時間にわたって安定した相関度であることを示す。また、2つのタグの平均相関度を重み付けすることも可能であるが、新しい相関度が高い重みを有する。これは図10に概略図示されている。図中、基準タグは"commuting"であり、タグ"nurenberg"の白色スペースが広がっていくことは、タグ"commuting"に対する"nurenberg"の相関関係が、ある時間にわたって減少していることを示す。他方、タグ"working"については、文字間の白色スペースが狭くなっていくことによって示されるように、ある時間にわたってタグ"commuting"との相関関係が増加している。

30

【0096】

また、y方向の個々の単線で示されるように、各タグについて、(例えば、より新しい相関値に、より高い重みが与えられる場合に、ある時間にわたって平均化された)全体的相関度が示される。タグ"commuting"との全体的相関関係は、タグ"transportation"と"working"とでほぼ同じであり、タグ"nuremberg"の場合より幾分か低いことが分かるであろう。

40

【0097】

白色間隔の相関分布は次式によって記述できる。

## 【数 8】

$$\vec{S}(\vec{C}) = [s_0, \dots, s_k] \text{ ここで}$$

$$\forall_{i=0}^k s_i = S_{\min} + \left(\frac{c_i}{C_{\max}}\right) \times (S_{\max} - S_{\min}),$$

$$(S_{\max} - S_{\min}) > 0,$$

$$\vec{C} = [c_0, \dots, c_k] \text{ 但し } \forall_{j=0}^k c_j, 0 \leq c_j \leq C_{\max}$$

10

ここで、

$s_i$  : タグ内の 2 セクション (例えば、文字) 間のスペース

$S_{\min}$  : タグ内の 2 セクション (例えば、文字) 間の最小 (デフォルト) スペース

$S_{\max}$  : タグ内の 2 セクション (例えば、文字) 間の最大スペース

$c_j$  : 関連タグについて測定された相関度

$C_{\max}$  : 最大相関度

である。

## 【0098】

20

平均重み付け相関度 A は次式によって記述できる。

## 【数 9】

$$A(\vec{C}, W) = \sum_{i=1}^k \left( \frac{c_i \times W}{i} \right) \text{ ここで}$$

$$\vec{C} = [c_1, \dots, c_k] \text{ 但し } \forall_{i=1}^k c_i, 0 \leq c_i \leq C_{\max},$$

$$W > 0$$

30

ここで、

A : 平均重み付け相関度

## 【数 10】

$\vec{C}$  : 所与のタグのある時間にわたる相関関係測定値ベクトル

$C_{\max}$  : 最大相関値

W : 重み係数

である。

40

## 【0099】

基準タグ ("commuting") が選択されると、固有数の相関タグが上方に配置され、基準タグと相関タグとの間の視覚的な距離増加または減少効果を強めるために僅かに回転させられる。そのような視覚化に基づく照会または閲覧は、例えば、前述の基準タグが再選択されるたびに、更に相関関係が小さい次グループタグの相関タグが同様に配置され、前に示された相関タグのグループと置き換わる。

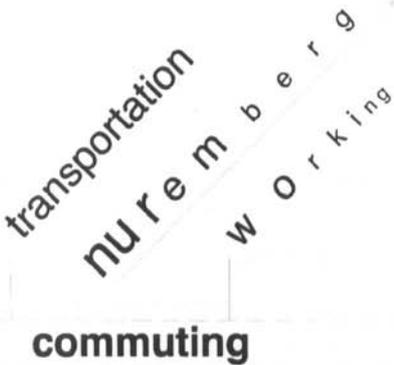
## 【0100】

下記では、上で説明された実施形態のうちのいくつかについて、対応するグラフィカルな視覚化表現と組み合わせて表示されるメタ情報を示すことによって簡単にまとめる。

## 【0101】

50

【表 2】

<p>タグ内の重要度分布</p>	<p>aaaaaaaaa</p>	
<p>タグ内の時間視覚化</p>	<p>C o m m u t i n g</p>	
<p>色分けされた相関重要度分布を備えた複数のシングルタグの類似重要度分布に基づいた一列状のグループ分け</p>	<p>traveler coffee munich iyout</p>	10
<p>固有タグの類似重要度分布に基づいた一列状のグループ分け。グループ形状は共通タグ判定基準を構成する。</p>	<p>traveler IYOUIT munich coffee</p>	
<p>固有タグの意味的または統計的類似性に基づく一列状のグループ分け。グループ形状は、ある時間にわたるグループの相関重要度を構成する。</p>	<p>laim munich nuremberg berlin</p>	20
<p>固有タグと対応基準タグとの間の相関度を視覚化するための色分け</p>		30
<p>相関度視覚化のための基準タグ選択。タグ内スペースの増加は相関度の減少を示し、スペースの減少は相関度の増加を示す。タグ内スペースの無変化は、ある時間にわたって安定した相関度であることを示す。</p>		40

【 0 1 0 2 】

今まで記載した実施形態では、フォントサイズ、彩度、スライス間の内部スペース、という3通りのタグのメタ情報視覚化法が主として用いられてきた。しかし、その時間的変

化を示すために個々のタグスライスのメタ情報を視覚化するための更に別の方法についても記載する。タグのスライスについて、このタグと関連のある特定量のメタ情報に対応する情報を図表的に視覚化するのに適した方法であれば、どのような方法も使用可能である。そのような一例が図 1 1 に示されている。図中、タグの微小要素（「スライス」）は、仮想階段上に配置されており、ある固有重要度測定値は、基線に対する実際の高さによって示されている。従って、重要度の高い要素は、重要度の低い要素よりも高く配置される。実際の高さ値の計算は、これまでのフォントサイズ計算と同様である。

【 0 1 0 3 】

ここで、

【数 1 1】

$\vec{L}$  : 所与のタグのある時間にわたる相関関係測定値ベクトル

$\vec{S}$  : ある時間にわたる所与のタグの重要度測定値ベクトル

$S_{max}$  : 最大重要度値

$H_{max}$  : 基線に対する最大高さ

$H_{min}$  : 基線に対する最小高さ

とする。

【 0 1 0 4 】

次式により、所与の範囲の重要度値が、結果として生じる高さベクトル

【数 1 2】

$\vec{H}$

によって示される固有タグの高さ値の個々の範囲にマッピングされる。

【数 1 3】

$\vec{H}(\vec{L}, \vec{S}) = [h_0, \dots, h_k]$  ここで

$$\forall_{i=0}^k h_i = H_{min} + \left( \frac{s_i}{S_{max}} \right) \times (H_{max} - H_{min}),$$

$$(H_{max} - H_{min}) > 0,$$

$$\vec{S} = [s_0, \dots, s_k] \text{ 但し } \forall_{j=0}^k s_j, 0 \leq s_j \leq S_{max},$$

$$\vec{L} = [l_0, \dots, l_k]$$

【 0 1 0 5 】

履歴上の重要度値測定の時間間隔が異なる場合もあるので、タグ内スペースによってこの時間的相違を示してもよい。一実施形態において、図 1 1 の左側に示されているように、小さいスペースは短い時間間隔に相当し、大きいスペースは長い時間間隔を意味する。

【 0 1 0 6 】

原則として、重要度、相関分布、または特定時間間隔を示すためにシングルタグに種々のフォントを使用することも可能であることは既に述べた（これは、Arial RegularおよびArial Narrowというフォントの使用を記載する例として与えられた先行する実施形態で説明された）。

【 0 1 0 7 】

10

20

30

40

50

しかし、これに加え、タグ"docomo"について図12に示されているように、特殊なタイプのフォントを使用できる。そのようなフォントの場合、文字内でさえもサイズを変更できる。そのようなフォントは、図12から分かるように、視覚化されるべき重要度値に従ってそのサイズを変えられる各文字の個々のスライスを提供することによって構成できる。

【0108】

さらに一般的に言うと、一実施形態において、メタ情報の視覚化は、メタ情報の量に基づいてタグの個々のスライスの視覚化パラメータを変更することによって実行され、この視覚化パラメータは、フォントスライスがそのようなメタ情報量を示すことができるように選ばれる。

【0109】

上記各実施形態は、ハードウェア、ソフトウェア、またはハードウェアとソフトウェアとの組合せによって実施できることが当業者には分かるであろう。本発明の実施形態に関連して記載されたモジュールおよび機能は、本発明の実施形態に関連して説明された方法に従って動作するように適切にプログラミングされたマイクロプロセッサまたはコンピュータによって、全体的または部分的に実施できる。本発明の実施形態を実施する装置は、例えば、計算装置または携帯電話または本発明の実施形態で記載された委任された処理 (delegated transaction) を実行できるように適切にプログラミングされた任意の携帯デバイス等を含むことができる。

【0110】

本発明の実施形態によれば、データ媒体に格納されるか、または、記録媒体もしくは伝送リンクなど一定の物理的手段によって具現化された他の手段に格納される、コンピュータ上で実行されると上記本発明の実施形態に従ってコンピュータを作動させることができるコンピュータプログラムが提供される。

【図2】

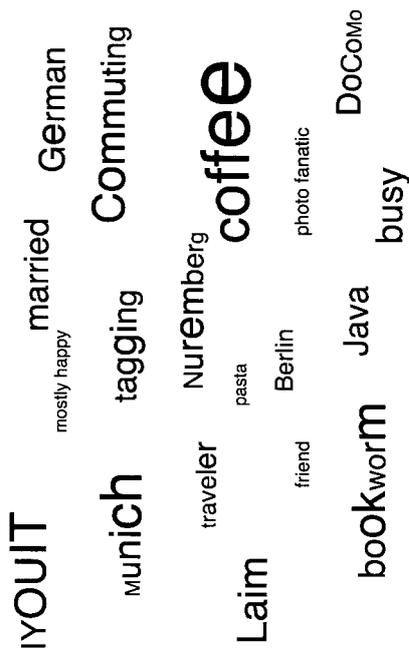


Fig. 2

【図3】

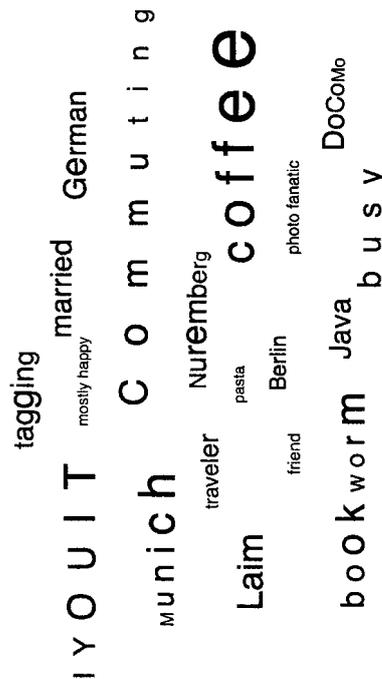


Fig. 3

【 図 4 】

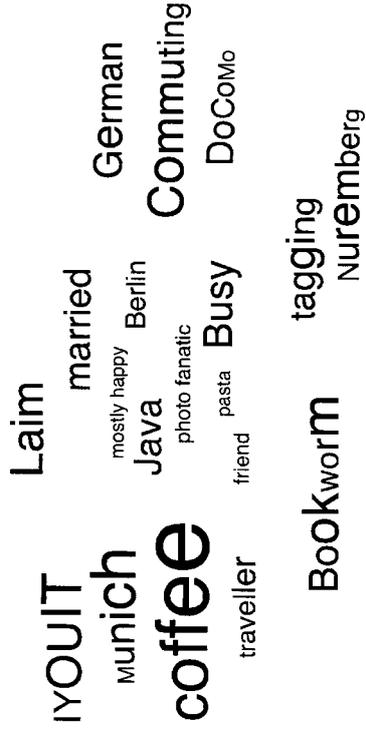


Fig. 4

【 図 5 】

...	Laim	...
IYOUIT munic coffee traveller	married Berlin Java photo fanatic friend pasta Busy	German Commuting DoCoMo tagging Nuremberg
Bookworm	...	

Fig. 5

【 図 6 】



Fig. 6

【 図 7 】



Fig. 7

【 図 8 】

aim munich nuremberg berlin

Fig. 8

【 図 9 】

Commuting  
Nuremberg  
coffee  
Bookworm  
Busy  
DoCoMo

Fig. 9

【 図 10 】

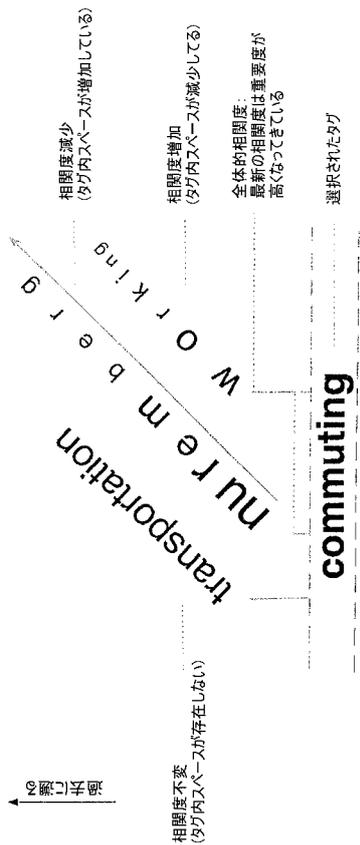


Fig. 10

【 図 11 】

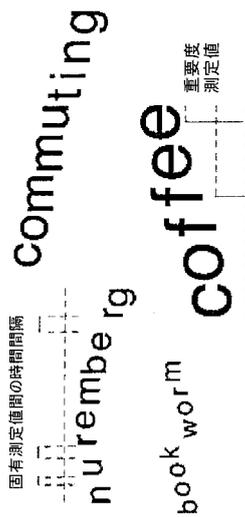


Fig. 11

【 図 1 2 】

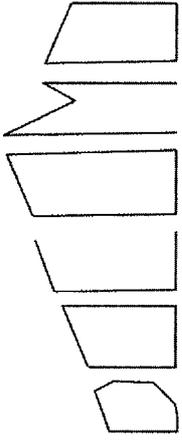


Fig. 12

【 1 】

**africa** amsterdam animals april architecture art asia australia baby band barcelona **beach**  
 berlin bird **birthday** black blackandwhite blue boston bw **california** cameraphone camping  
 canada canon car cat chicago china christmas church city clouds color concert  
 cute dance day de dog england europe fall **family** festival film florida flower  
**flowers** food france friends fun garden geotagged germany girl girls graffiti  
 green halloween hawaii hiking holiday home honeymoon house india ireland island **italy**  
**japan** july kids la lake landscape light live london macro me mexico mountain  
 mountains museum music nature new newyork newyorkcity newzealand night nikon  
 nyc ocean paris park party people photo photos portrait red river rock rome san  
 sanfrancisco scotland sea seattle show sky snow spain spring street summer  
 sun sunset sydney taiwan texas thailand tokyo toronto **travel** tree trees trip uk urban  
**usa vacation** vancouver washington water **wedding** white winter yellow york  
 zoo

**Fig. 1**

## フロントページの続き

(74)代理人 100125036

弁理士 深川 英里

(74)代理人 100142996

弁理士 森本 聡二

(74)代理人 100154298

弁理士 角田 恭子

(74)代理人 100156443

弁理士 松崎 隆

(74)代理人 100162330

弁理士 広瀬 幹規

(72)発明者 セバスティアン・ベーム

ドイツ連邦共和国, 8 1 6 7 1 ミュンヘン, ヘヒトゼーシュトラッセ 1 4

(72)発明者 マルコ・ルター

ドイツ連邦共和国, 8 2 2 5 6 フュルステンフェルトブルック, カプツィーナー・シュトラッセ  
1 8

(72)発明者 マティアス・ヴァーグナー

ドイツ連邦共和国, 8 0 9 9 9 ミュンヘン, ガンツェンミュラーシュトラッセ 3 3

Fターム(参考) 5E501 AA01 AB15 AC15 BA05 BA08 CA02 CB09 EA05 EA13 EB05

FA13 FA23 FA43 FA45 FB25 FB26 FB28 FB43

【外国語明細書】

## **1. TITLE OF INVENTION**

### **METHOD AND APPARATUS FOR VISUALIZING A TAG CLOUD**

## **2. DETAILED EXPLANATION OF THE INVENTION**

### **5 FIELD OF INVENTION**

The present invention relates to a method and an apparatus for visualizing a tag cloud.

### **BACKGROUND OF THE INVENTION**

10

Tag clouds have become a popular way to represent (unstructured) data by assigning significance values (based on the frequency within the given dataset) to tags within multiple, congregated tags. A so-called tag cloud can be seen as an aggregation of numerous tags with or without a certain order, for instance tags within a cloud are often sorted alphabetically. The significance of a tag is indicated by the font size used for the respective tag, so a lower significance is indicated by a smaller font size for the given tag. Here, a tag is a keyword that refers to a certain piece of information, therefore enabling a simple classification and categorization. In the following, the term tag will be used as an umbrella term for any single piece of information that can be represented within a cloud formation and that may also be defined by several attributes not essentially visible in the first place. Today, tag clouds are mostly used for finding popular data items that have been previously marked (tagged) with a certain keyword for later reference and to indicate popular data items within a larger dataset based on their frequency within the dataset. An example of a conventional tag cloud in which the significance of a tag is indicated by the font size used for the respective tag is shown in Fig. 1.

15

20

25

Despite their popularity, tag clouds are rather limited with respect to the amount of meta-data (descriptions about the actual information) that can be visualized and do only provide limited means of navigating through the represented dataset. One major  
5 drawback of common tag cloud visualization techniques is the fact that only a snapshot of a given dataset for a certain moment in time can be displayed, thus only the latest sample of the dataset can be represented. As a result, current tag cloud visualizations do not allow for drawing any conclusions about the significance of one or multiple tags *over time*. Hence, the application of tag clouds for the representation of dynamic and historic  
10 information is limited since only popular items at a certain moment in time can be highlighted. The significance as well as the correlation distribution within a dataset over time is so far not considered.

Being mostly used for the visualization and representation of unstructured information,  
15 each tag within the cloud is treated and represented independently of other tags of the same cloud. This means that even simple correlations between tags that indicate basic co-occurrence measurements or potentially more advanced semantic similarities cannot be expressed and therefore do not influence the visual significance of single tags. Current tag clouds clearly lack the possibility to indicate a certain correlation or connectedness  
20 between tags of the same cloud, and thus limit the scope of tag clouds to visualize implicit information explicitly and only provide very basic means of interacting with the represented data.

So far there is no solution known that allows for visualizing the temporal significance and  
25 correlation distribution within single tags or groups of tags in a 2D tag cloud for identifying

temporal coherences that may also be used to navigate through historic samples of the dataset and to restrict queries based on the significance and correlation distribution over time. Existing tag cloud representations are basically limited to the visualization of a single tags' frequency within a given dataset at a certain moment in time. As a result, the application of tag clouds for navigating through a given dataset is very limited.

In "Yusef Hassan-Montero, Víctor Herrero-Solana, Improving Tag Clouds as Visual Information Retrieval Interfaces. International Conference on Multidisciplinary Information Sciences & Technologies, 2006" the authors propose clustering techniques to group tags within a cloud formation that have been identified as "similar", based on relative co-occurrence measurements. Even though this approach does help to visually identify similar tags, no temporal aspects with respect to the tag's significance distribution as well as their correlation distribution are considered. In addition, the proposed visualization technique does not improve the functionality of the tag cloud with respect to navigation through the given dataset. Also, the grouping algorithm is only taking co-occurrence measurements into account, rather than a common trend or the correlation distribution over time.

In "Owen Kaser and Daniel Lemire, Tag-Cloud Drawing: Algorithms for Cloud Visualization, Tagging and Metadata for Social Information Organization. World Wide Web Conference, 2007", the authors focus on algorithms for improving existing tag cloud visualization techniques with respect to optimizing the amount and placement of white space within the cloud. The authors also propose to utilize the font spacing attribute to describe the visual weight and importance of single tags, however, the inner-tag space remains constant and does not encode a temporal significance or correlation distribution.

Grouping of tags based on co-occurrence measurements is applied within the cloud.

An attempt to visualize tags over time is described in "Micah Dubinko, Ravi Kumar, Joseph Magnani, Visualizing Tags over Time. World Wide Web Conference, 2007". The authors follow an animation-based approach, which is why their approach cannot be applied in a static 2D cloud. The principle idea is to present a selection of tags that are meant to represent a pre-defined time interval that has been selected by means of a time-slider. However, this approach is concerned with characterizing a certain time period (by means of a selection of tags) rather than describing tags over time. Furthermore, no query restriction through the interaction with the tags can be achieved. The selection of tags is only based on their frequency within the given time interval.

In "Blake Shaw, Utilizing Folksonomy: Similarity Metadata from the Del.icio.us System, <http://www.metablake.com/webfolk/web-project.pdf>", the author defines a similarity measurement that goes beyond a basic co-occurrence measurement. The similarity measurement considers information about the actual tag contributor and his tag selection in the past. As a result, a distance-based visualization of similar tags is proposed. Again, no temporal aspects are considered with respect to a tag's significance as well as correlation distribution over time, neither for the actual visualization of tags nor for the proposed interaction pattern with the cloud. Therefore, no grouping of tags based on those temporal aspects (their trend or correlation distribution over time) can be applied.

## **SUMMARY OF THE INVENTION**

According to one embodiment there is provided a method of visualizing a tag cloud of individual tags, wherein each is a certain piece of information and is formed of one or more characters, and wherein for a tag metainformation has been measured over a certain time interval so that there exists a vector of metainformation measurements for said tag, each component of said vector representing a certain metainformation measurement at a certain time, said method comprising:

- 5 dividing a tag into a plurality of slices;
- mapping said vector of metainformation measurements of said tag over time such that each slice of said tag corresponds to one or more of said metainformation measurements;
- 10 visualizing the temporal change of said metainformation measurements for said tag by changing for each of said slices one of the following parameters depending on the corresponding one or more metainformation measurements:
  - the font size,
  - the color saturation;
  - 15 the inner space between said slices.

In this way metainformation which has been recorded over time for a certain tag may be visualized in the tag cloud so that the resulting tag cloud offers temporal information about how measured (or inputted) data change over time. This can then be used for an interface

20 which enables querying or browsing said data based on such temporal aspects and thereby opens a new dimension of how to interact with and handle data displayed by tag clouds for the purpose of querying and browsing.

According to one embodiment wherein said metainformation is one of the following:

- 25 the significance of a tag,
- the correlation significance between tags.

These two types of information are of particular interest for a user when browsing a dataset or when querying data, and therefore it is advantageous if one or both of them ca

be displayed.

According to one embodiment the total period of time which is mapped onto a certain tag is graphically represented by the total length of said tag.

5

this gives a user an impression about the temporal coverage of a certain time as indicated in the tag cloud.

According to one embodiment the method comprises:

- 10 identifying for each tag to which temporal pattern of a plurality of predefined temporal pattern it belongs, and  
grouping the tags together which belong to the same pattern.

15 This makes it easier for a user to identify those tags which show the same temporal pattern, and he may select one of them. Moreover, such a grouping can also be used for browsing or querying, e.g. by selecting a certain group about the tags of which then further information may be retrieved

20 According to one embodiment a plurality of tags which belonging to the same group of tags are selected by a similarity criterion, or arbitrarily by a user, and wherein for the tags of said group metainformation is displayed by one or more of the following:  
a change in font size;  
25 a change in color saturation.

Grouping claims together can make sense, either by selecting them by a user or by a similarity criterion (such as similar shape of significance over time), and then for the tags of the group metainformation may be display by means of change in font size or color

saturation

According to one embodiment a plurality of tags which form a group is aligned in a row and said group is divided into slices, each slice representing a certain time period,  
5 wherein the vector of metainformation to be visualized is representing the correlation significance between said elements of said group over time, said vector of metainformation being mapped onto said slices of said group of tags such that the visual representation of said group of tags illustrates the change of the correlation between the tags of said group over time based on a change of font size or color saturation.

10

This makes it possible to visually display the temporal change of the overall correlation between the elements of the group of tags over time.

In one embodiment thereby the ordering of the tags in said group is chosen such that the  
15 individual tag having the highest individual significance is located where the correlation significance is highest (which may mean at the location where the font size is largest or the color saturation is highest).

According to one embodiment the method comprises:

20 selecting a group of tags based on a similarity criterion and aligning them in a row and said group being divided into slices, wherein the overall shape of the group is based on the temporal change of the metainformation which corresponds to the similarity criterion, and/or  
with the tags of the group being ordered such that the tag of the group which has the  
25 highest individual significance is located at the position with the highest font size or color saturation and the other tags are ordered according to their correlation with the tag having the highest significance.

By the overall shape of the group being dictated by the temporal change of the similarity  
30 criterion it becomes possible to readily see how the similarity criterion (which may e.g. a

certain pattern of the individual tag significance, such as “rising” or “falling”) looks like. Moreover, by the ordering of the tags in the row according to correlation with the tag having the highest significance furthermore correlation information can be visualized.

- 5 According to one embodiment a group of tags has been selected based on a similarity criterion according to which the members of said group have a certain similarity, and the overall correlation significance of the correlation between the group members over time is visualized by
- a changing font size which changes in accordance with the correlation significance
- 10 change over time; or
- a color saturation change which changes in accordance with the correlation significance change over time.

In this way the overall correlation between the members of the group may be visualized.

15

According to one embodiment the method comprises:

- selecting one tag as reference tag;
- graphically visualizing the correlation of other tags with said selected tag by displaying said other tags such that the degree of correlation with said reference tag is expressed by
- 20 the font size of said other tags; or
- the color saturation of said other tags.

In this way a correlation between a reference tag and other tags may be visualized.

- 25 According to one embodiment the inner significance of a tag over time is expressed by one of its font size or color saturation; and
- the correlation significance of a tag with said selected tag over time is represented by the other one of said font size or said color saturation.
- In this way the significance of a tag itself and furthermore the significance of its correlation

with the reference tag both may be visualized over time.

According to one embodiment the method comprises:

- 5 a reference tag being selected and other tags are simultaneously displayed in a manner which indicates their respective correlation with said selected tag over time, wherein said correlation is indicated by
- the space between slices or characters of a tag are changing in accordance with the change of said correlation, or
- the color saturation changing in accordance with the change of said correlation.

10

According to one embodiment the method comprises:

in addition to visualizing the correlation of a tag with said reference tag over time, further visualizing the overall correlation of said tag with said reference tag.

- 15 This enables also the visualization of the overall correlation of a certain tag with the reference tag in addition to the correlation as it changes over time. The overall correlation may e.g. be displayed by the extension of a single line located close to the tag.

According to one embodiment the visualization of a tag cloud enables the user to query  
20 or browse context information as represented by said tag cloud based on the metainformation represented in said tag cloud, said metainformation being one or more of the following:

- temporal change of a tag significance over time,
- temporal change of a correlation significance between tags over time,
- 25 any metainformation about the tags of said tag cloud as graphically visualized by said tag cloud.

According to one embodiment there is provided an apparatus for visualizing a tag cloud of individual tags, wherein each is a certain piece of information and is formed of one or

more characters, and wherein for a tag metainformation has been measured over a certain time interval so that there exists a vector of metainformation measurements for said tag, each component of said vector representing a certain metainformation measurement, said apparatus comprising:

5 means for dividing a tag into a plurality of slices;

means for mapping said vector of metainformation measurements of said tag over time such that each slice of said tag corresponds to one or more of said metainformation measurements;

10 means for visualizing the temporal change of said metainformation measurements for said tag by changing for each of said slices one of the following parameters depending on the corresponding one or more metainformation measurements:

the font size,

the color saturation.

15 In this way an apparatus according to an embodiment of the invention may be implemented.

According to a further embodiment the apparatus comprises means for carrying out a method according to any of the embodiments of the invention.

20

According to one embodiment there is provided a computer program comprising computer program code which when being executed on a computer enables said computer to carry out a method according to any of the embodiments of the invention.

## 25 **DESCRIPTION OF THE DRAWINGS**

Fig. 1 illustrates a tag cloud as known in the art.

Figures 2 to 12 schematically illustrate tag cloud representations according to embodiments of the invention.

## DETAILED DESCRIPTION

The concept according to an embodiment of the invention considerably extends the basic  
5 concept of tag clouds for data visualization and navigation by taking temporal as well as  
correlation aspects into account. With the proposed visualization and interaction  
schemes, new ways to represent and query large datasets can be realized as will become  
apparent in the following.

10 In particular, it will be described how significant data items in large datasets can be  
visually described and utilized for browsing based on novel visualization techniques that  
encode additional information (metainformation) in tag clouds to highlight the significance  
of single as well as correlated tags over time.

15 The metainformation can be automatically recorded, e.g. by recording in regular intervals  
the data from a GPS sensor based on which position information is determined, and  
based on the position information there may then be automatically recorded a  
corresponding tag, such as e.g. the name of the city of the position, the name of the  
location (e.g. office), and the like. The tags may, however, also be based on  
20 metainformation which is inputted by a user and represents e.g. the user's context, such  
as his present mood (lazy, happy, sad, or the like) or his activities (sports, reading,  
walking). Such metainformation may be inputted by the user and may then be recorded  
together with the corresponding time. The term "measurement" therefore has to be  
understood that also the input of metainformation by a user at a certain time may be

regarded as a "measurement" of metainformation at a certain point in time.

All such metainformation which is either derived from user input or from automatic measurements may then result in a compilation of several kinds or types of metainformation, where the different kinds or types may be represented by corresponding tags the significance of which is recorded over time (e.g. by the number of occurrence within a certain time interval).

A thus collected pool of context data (or "measurement data") which represents context information over time may be the basis on which the visualization mechanisms of the present invention may be applied in order to enable the user to meaningfully view, browse and query this pool of information. Since the context information may be very different, and since it also contains temporal information, the present interfaces for displaying and querying such a pool of information are not suitable. Accordingly the embodiments of the invention as described in the following can help the user to access this data, to query it and to browse it in a manner which was not possible before.

A tool which may be used to collect the metainformation may be a mobile device such as a mobile phone, a smartphone, or the like. It may execute the automatic measurements (such as position measurements, or any other measurement of data which result from the sensors of the device) and may enable the user to input metainformation which then may be recorded together with the corresponding time. The mobile device may then also perform a visualization method as described in the embodiments of the invention by executing a computer program which performs the corresponding functions, and it may

also allow the user to query and browse through the metainformation which results from the data collection by the "measurements" which record metainformation over time, the metainformation corresponding to and being represented by corresponding tags.

5 The proposed techniques in one embodiment allow for describing the significance distribution of a tag over time in changing its font characteristics (e.g. by using different font sizes) within the representation of one tag. So in case the significance of a tag has decreased significantly, the font size of the distinct tag decreases, too, and vice versa. As a result, the tendency of the observed significance of multiple tags can be visualized in  
10 one (static) cloud, and thus enables a better identification and selection of tags not only based on their current significance, but their significance distribution over time. In addition, being able to visually identify trends of single data items or groups of items, it becomes possible to not only better understand the provided information but also to browse to historic samples of the data as recorded in the past. By selecting distinct parts of a single  
15 tag's representation, historic samples of this data item can be revealed. Likewise, through visualizing a tag's significance distribution over time, a trend-based browsing through the given dataset becomes possible. For example, one can choose a tag which has a desired trend over time (e.g. increasing significance) and can search for other tags (and their related information) which have the same trend over time. Therefore the visualization  
20 technique of the embodiments of the invention not only visualizes data which has been recorded over time, but it also enables a novel interface to query and browse through such data.

Today, the significance of tags in a cloud is so far only based on their frequency within a  
25 given dataset. In some cases, however, the frequency alone is not a suitable indicator for

the significance and therefore importance of a tag. Instead, the significance of a tag is also heavily depending on the correlation or connection with other tags in the same cloud as well as the significance of the correlation over time. The correlation degree of two or more tags can be based on co-occurrence measurements and may also indicate a semantic or statistical similarity (closeness) of tags. Existing tag cloud representations do not allow for visualizing the degree of correlation and therefore lack a suitable differentiating factor between tags in one cloud. As a result, navigation through the dataset and restricting a given query along the lines of correlated tags is not possible.

10 The visualization and interaction techniques according to embodiments of the invention tackle this constraint by visualizing the correlation factor between tags with either a distinct grouping, color-coding or distance indication. The correlation significance of a group of tags over time can be visualized by adjusting the entire group's shape to match its correlation significance. The color-coding paradigm allows for selecting a certain tag, hereupon highlighted by a reference color with 100% saturation. In one embodiment all correlated tags are accentuated in the same color tone, but based on the correlation degree, the saturation percentage is chosen whereas a higher degree of correlation is equivalent to a higher color saturation. In one embodiment, by adjusting inner-tag white spaces, the correlation degree over time can be represented. In addition, in a further embodiment recent correlation degrees may have a higher emphasis in the overall correlation degree of two given tags.

With the proposed visualization and interaction techniques, new ways to navigate and browse through large datasets become possible through the generation of advanced tag cloud representations. Queries can be restricted based on the significance distribution of single items or groups of tags, and may also allow for browsing through correlated items

and provide means to navigate to historic samples of a given dataset. Queries may thus be restricted based on temporal conditions, certain correlation dependencies or similar trends that were identified over time.

5 With the proposed visualization and interaction paradigms, significant items in datasets can be highlighted not only based on their current significance, but also based on their significance distribution over time and the significance of their correlation with other tags over time. This way, only recently popular (short-term) tags can be isolated and significant tags over a longer period of time and with a high degree of correlation gain more visual  
10 weight. In addition, the proposed techniques allow for a better selection of single data items or groups of items based on their significance and correlation distribution over time and allow for querying historic data samples that have been recorded in the past, provide means to isolate groups of tags with a similar significance and correlation trend and to browse through the given dataset based on the correlation degree between tags.

15

Now embodiments of the invention will be described in somewhat more detail.

At first an embodiment where an Inner-tag significance distribution is visualized will be explained.

20 A single's tags significance distribution is defined as the number of significance measurements that were recorded within a certain time period. With it, a certain tendency of a tag's importance or interestingness can be described over time.

The proposed visualization and interaction techniques considerably extend existing tag

cloud representations, enable the visualization of additional meta-information in static clouds and allow for novel interaction and query restriction patterns. Metadata allows for identifying and highlighting the significance distribution of single tags within the cloud and the correlations between two or more tags over time.

5

With the proposed visualization and interaction paradigm, the significance of a tag is not only based on its frequency within a given dataset at a certain moment in time (one single value), but may consist of a number of significance levels recorded over time. The resulting significance distribution holds valuable information about the significance of a tag during a certain time period, so whether or not a tag's popularity has been decreasing, increasing or may have remained the same. With the proposed visualization paradigm, this tendency is made visible within a single, static tag cloud, as depicted in Figure 2. In the example below, the increasing significance of the tag "coffee" is indicated by its increasing font size, whereas the tag "commuting" is decreasing in terms of font size and its corresponding significance over time.

10  
15

The actual font sizes used for a given tag and its significance distribution can be computed by assuming that the number of distinct sections (a section may e.g. a letter or a smaller part of the tag, e.g. a slice of a letter, or any part into which a tag may be divided) of a given tag also restricts the number of significance measurements considered.

20

With:

$\vec{L}$ : Vector of distinct sections of a given tag

$\vec{S}$ : Vector of significance measurements of a given tag over time

25  $S_{\max}$ : Maximum significance value

$F_{\max}$ : Maximum font size

$F_{\min}$ : Minimum font size

The formula below maps the given range of significance values to the respective font sizes of a distinct tag, described by the resulting font size vector  $\vec{F}$ .

$$\vec{F}(\vec{L}, \vec{S}) = [f_0, \dots, f_k] \text{ with}$$

$$\forall_{i=0}^k f_i = F_{\min} + \left( \frac{s_i}{S_{\max}} \right) \times (F_{\max} - F_{\min}),$$

$$(F_{\max} - F_{\min}) > 0,$$

$$\vec{S} = [s_0, \dots, s_k] \text{ with } \forall_{j=0}^k s_j \cdot 0 \leq s_j \leq S_{\max},$$

$$\vec{L} = [l_0, \dots, l_k]$$

10

Through the visual representation of a tag's significance distribution, temporal query restrictions can be defined. The horizontal display of a tag represents the observed time interval, whereas distinct sections of a tag represent a temporal subset of that time interval. The number of time slices (=the sections of the tag) that can be identified as well as the actual number of significance measurements may depend on the overall size of a certain tag. Selecting a certain time slice can be seen as a temporal query restriction, whose result set corresponds to the chosen time interval. This way, single pieces of information in large datasets can be retrieved through further narrowing down the observed time interval. A slice of a tag may correspond to a single character of the tag or it may correspond to a part (a "slice") of a character. In first case the significance indication varies from character to character, but remains the same within a single

20

character. In the latter case the significance indication may even change with a character, in other words the font size of an individual character may even change within the character.

If the number of sections of the tag and the number of temporal measurements are not the same, in other words if the vector of distinct sections of a given tag and the vector of significance measurements of a given tag over time have a different number of components, then there may be performed a matching such that a plurality of components of one vector are mapped onto a single component of the other vector. This can e.g. be done by choosing the average significance value of the plurality of significances, if several significance components are to be mapped onto a single tag section.

The time interval which is represented by the tags of the tag cloud may be predetermined (e.g. one day, one week, one month or the like), or it may be selected by the user. The user may e.g. choose the time interval which he is most interested in, and based on the collected meta-information the tag cloud which visualizes the temporal development of significance data may be calculated and displayed.

According to one embodiment, since not all tags that form a tag cloud necessarily represent the same observed time interval (e.g. because not for all of them there is enough meta-information data to cover the whole interval selected by the user for being visualized), according to one embodiment an additional indicator for the relative time period considered is used. In this embodiment the total (horizontal) length of a tag is a measure for the total time period over which the significance measurement extends. The length variation to represent different time intervals by different tags can e.g. be achieved by varying the inner-tag white spaces. As shown in Figure 3, the variation of inner-tag white

spaces increases its horizontal extension, which corresponds to a relatively longer time period. In the example shown in Figure 3, the tag “commuting” covers a longer time period than “nuremberg”, for instance.

This additional inner-tag white space or font spacing can be approximated by:

$$5 \quad \bar{D}(\vec{T}) = [d_0, \dots, d_k] \text{ with}$$

$$\forall_{i=0}^k d_i = D_{\min} + \left(\frac{t_i}{T_{\max}}\right) \times (D_{\max} - D_{\min}),$$

$$(D_{\max} - D_{\min}) > 0,$$

$$\vec{T} = [t_0, \dots, t_k] \text{ with } \forall_{j=0}^k t_j \cdot 0 \leq t_j \leq T_{\max}$$

With:

$d_i$  : Space or distance in between two inner-tag's sections (e.g. letters)

10  $D_{\min}$  : Minimum (default) space between two inner-tag's sections (e.g. letters)

$D_{\max}$  : Maximum space between two inner-tag's sections (e.g. letters)

$t_j$  : Observed time interval for the concerned tag

$T_{\max}$  : Maximum observed time interval for a tag in a given cloud

15

Alternatively to a variation of the inner-tag white spaces, different font types or variants can be used. Tags rendered in a more condensed font (e.g. Arial Narrow) represent a shorter time interval, regular or wider fonts represent larger time periods (e.g. Arial Regular).

The visualization of a tag cloud as described before enables a user to navigate through context information as represented by such a tag cloud based on temporal information. For example a user may select a tag which with respect to its temporal change represents a rise, so that it has become more significant recently, and thereby the user may choose context information for navigating through a dataset based on temporal information. Another possibility is to select not a single tag but a part of the tag which then represents a certain time interval being a subset of the whole time interval represented by this tag. Such a selection could e.g. be made by drawing a box around the selected part of the tag using the mouse or the cursor in a well known manner, and the thus selected time interval is then applied as a filter or selection criterion when obtaining further information like other information related to the selected tag and time. In one embodiment the selection of a subset in time may lead to a recomputation of the whole tag cloud for the thus selected subset in time, thereby effectively zooming into the tag cloud..

15

In a relatively simple case, a distinct section of the tag (e.g. one single letter) could be selected (e.g. by clicking on it), which already represents the aforementioned temporal selection and query restriction. In addition, in one embodiment selecting more than one parts of the tag could be accomplished in a similar way as marking a certain text passage in word processing. The selection could then be confirmed by a mouse-click.

20

In contrast to selecting distinct parts of a tag to restrict the resulting query with respect to the observed time period, the selection of the space in-between a tag according to one embodiment has the opposite effect. In this case, according to one embodiment the considered time period will be extended, resulting in a re-computation of each single tag

25

within the cloud.

Now a further embodiment will be described which makes use of a tendency-based grouping of tags.

- 5 Having computed the significance distribution of a single tag over time as exemplified in Figure 2, distinct tags can not only be grouped based on their (semantic) similarity or based on simple co-occurrence measurements but can furthermore be grouped based on similar significance trends. In this respect, several types of trends can be identified as listed below (amongst others):

10

**bookworm**

Increasing significance with:

$$\forall_{i>0}^l s_i < s_{i+1} < s_l \text{ (s}_i\text{: distinct significance measurement)}$$

**bookworm**

Decreasing significance with:

$$\forall_{i>0}^l s_i > s_{i+1} > s_l \text{ (s}_i\text{: distinct significance measurement)}$$

15

**bookworm**

Recurring, increasing significance with:

$$\forall_{i=0}^{k-1} s_i < s_{i+1} \wedge \forall_{j=k+1}^l s_j < s_{j+1} \wedge s_k > s_{k+1}$$

**bookworm**

Recurring, decreasing significance with:

$$\forall_{i=0}^{k-1} s_i > s_{i+1} \wedge \forall_{j=k+1}^l s_j > s_{j+1} \wedge s_k < s_{k+1}$$

**bookworm**

Decreasing and increasing significance with:

20

$$\forall_{i=0}^{k-1} s_i > s_{i+1} \wedge \forall_{j=k+1}^l s_j < s_{j+1} \wedge s_k = s_{k+1}$$

**bookworm**

Increasing and decreasing significance with:

$$\forall_{i=0}^{k-1} s_i < s_{i+1} \wedge \forall_{j=k+1}^l s_j > s_{j+1} \wedge s_k = s_{k+1}$$

A tag cloud that has been reformatted by having identified groups of tags with similar significance distributions is shown in Figure 4.

- 5 Having identified certain groups within a tag cloud formation based on similar significance trends, a group based query restriction according to one embodiment reconsiders the given result-set with respect to the chosen significance trend only.

The layout of the resulting tag cloud formation according to one embodiment depends on  
10 the following paradigm: Each group of tags can in principle be seen as a detached cloud. The largest group of tags is placed in the middle of the resulting (combined) tag cloud, all subsequent groups are placed around the centered group. In the example shown in Figure 4, the cloud containing tags like "laim" and "busy", for instance, has been identified as the largest cloud and is therefore placed in the middle of the resulting combined tag  
15 cloud. The alignment is based on a matrix where the number of fields must be larger than the total number of identified trends (and such groups of tags). This way, an additional visual weight is put to the largest and therefore most important group of tags, which in turn facilitates a quicker navigation, selection and identification of single tags. Fig. 5 illustrates the matrix structure of the tag cloud shown in Fig. 4.

20

In the following there will be described a embodiment which makes use of color coded correlation significance distribution

In addition to the temporal significance distribution of a tag, its correlation with other tags of the same cloud over time is valuable metadata, too. The temporal correlation  
25 distribution can be defined as the number of correlation measurements of a certain group

of tags as recorded during a specified time interval. The correlation as such can be based on co-occurrence measurements, a semantic dependency or any connection between those concepts that has been identified either with respect to the time, the physical space, or the common presence of another data item not present within the group.

5

Contemplating identified groups of tags that have a common significance distribution over time allows for visualizing the group's correlation significance in a one-line formation, in which the x-axis can be seen as a timescale with the most recent measurements on the left, and the oldest measurement on the right (or otherwise round). A number of those tags grouped in one line again form a 2D tag cloud. Each tag within the group still holds its specific information about its own significance distribution, while the group's correlation significance over time is indicated by a fading color schema with 100% color saturation representing the highest correlation significance. Such a tag cloud may then look like as illustrated in Fig. 6.

10

15

The tags belonging to the group of Fig. 6 may not only be selected based on a similar temporal significance distribution, the tags may just be arbitrarily selected, e.g. by the user, and then the correlation between the tags over time may be displayed in the manner as shown in Fig. 6, namely by the color saturation varying over the x-axis which corresponds to the time.

20

It should further be mentioned that while different selection criteria for selecting the members of a group are mentioned in this and other embodiments, all these selection criteria could of course also be combined in any combination. E.g. the tags could also be selected based on their semantic similarity as described in another embodiment.

25

Alternatively or additionally also always an arbitrary selection by the user is possible in one embodiment.

The color saturation that represents a groups' correlation distribution can be described

5 by:

$$\forall_{i=0}^k x_i = X_{\min} + \left(\frac{c_i}{C_{\max}}\right) \times (X_{\max} - X_{\min}),$$

$$(X_{\max} - X_{\min}) > 0,$$

$$\vec{C} = [c_0, \dots, c_k] \text{ with } \forall_{j=0}^k c_j, 0 \leq c_j \leq C_{\max}$$

With:

$x_i$ : Color saturation for the respective section within a group of tags

10  $X_{\max}$ : Maximal color saturation

$X_{\min}$ : Minimum color saturation

$c_i$ : Given correlation significance sample

$C_{\max}$ : Maximum correlation significance of the group

15 According to one embodiment also the ordering of tags within the group follows a certain paradigm. The most significant single tag is placed where the group's correlation significance is at its peak (in Figure 6, "coffee" has been identified as the most significant single tag, thus placed at the second position from left where the group's maximum correlation significance has been measured).

20

Group criterion: Similar significance distribution (or arbitrary tag selection)

Tag shape: Respective tag's significance distribution

Tag placement: Place of the most significant single tag corresponds with the place of max

correlation significance of the group formation

Group shape: -

Color-coding: Fading color indicates the group's correlation significance over time

Interaction: Query restriction based on time period and common group criteria

5 (similar significance distribution)

Now an example of how this visualization pattern may be used for navigation purposes will be described. For example, to further refine a given query and to diminish the result-set, distinct parts of the group can be selected. Hereby, the time period in which data items must have been recorded will be restricted to correspond to the selected region of the displayed correlation significance. Since not only a time period has been identified but also a certain group has been selected within the cloud, the resulting query will in one embodiment will also be defined by the common group criteria, in the example above a similar significance distribution over time.

15

In the following there will be described a embodiment which makes use of group-based significance distribution.

In this embodiment the actual correlation between the tags of the group as well as the group's common criteria are of subsequent interest. Focusing on the group as such, its common criteria as well as the correlation between the tags within the group can be visualized as illustrated in Figure 7.

Here, the group's common criteria (a similar significance distribution) forms the whole group's shape, again aligned in one line, while the most significant tag of the group (in this case "coffee") is identified and subsequently visualized as the most prominent tag (in

25

terms of font size in Fig. 7, could also be color saturation if that is chosen to express significance). The ordering of the given tags within the group formation is based on the correlation *degree* of a certain tag with respect to the most significant tag of the group. The correlation degree describes the dependency between two distinct tags within a  
5 group. In the example above, the tag "munich" has a higher correlation degree with "coffee" than "traveler".

The shape rendering can be described mathematically analog to a single tags' significance distribution as formalized in the first example, whereas the group of tags is  
10 hereby treated as one joined, single tag.

Group criterion: Similar significance distribution

Tag shape: Determines (and thus forms) the group's shape

Tag placement: Most significant single tag is placed at group's max significance;  
15 remaining tags are ordered based on the correlation degree with respect to the most significant single tag

Group shape: Common criterion (significance distribution over time) defines the group's shape

Color-coding: -

20 Interaction: Query restriction based on time period and common group criteria (similar significance distribution)

Now there will be described an embodiment which makes use of a group-shaped correlation significance distribution.

25 Characterizing the correlation of several tags over time, describes the significance of the

correlation of the group as such. Thus all tags within the group have been selected based on a certain correlation (e.g. based on a semantic similarity), rather than a similar significance distribution as discussed in the previous examples. The proposed visualization paradigm then depicts the whole group's correlation significance over time  
 5 whereas each distinct tag's significance distribution is not the prime intention any more, as shown in Figure 8.

The example that has been selected and visualized in Figure 8 describes a group of similar tags (all tags represent spatial concepts, the similarity criterion therefore here is  
 10 "tags representing a spatial concept") and the group's correlation significance over time. Here, the significance of the above shown group has been first increasing, then followed by a steady decrease of the correlated significance. The shape of each single tag complies with the group's shape as such, which leads to a more prominent visualization of the entire group and its correlation significance over time. Again, the ordering of the tags  
 15 within the group follows a certain principle. That is, the most significant single tag takes over the place where the highest significance value(s) are present. As a result, "munich" as the most significant tag of the group has been placed within the first part of the cloud formation, followed by "nuremberg", "laim" and "berlin".

20 The distribution of the correlation significance of the group can be described analog to the inner-tag significance distribution (as explained before), whereas the resulting Font-size vector is not described by the respective significance measurements but in fact determined by the group's correlation measurements.

25  $\vec{F}(\vec{L}, \vec{C}) = [f_0, \dots, f_k]$  with

$$\forall_{i=0}^k f_i = F_{\min} + \left(\frac{c_i}{C_{\max}}\right) \times (F_{\max} - F_{\min}),$$

$$(F_{\max} - F_{\min}) > 0,$$

$$\vec{C} = [c_0, \dots, c_k] \text{ with } \forall_{j=0}^k c_j \cdot 0 \leq c_j \leq C_{\max},$$

$$\vec{L} = [l_0, \dots, l_k]$$

With:

- 5  $\vec{L}$ : Vector of distinct sections of the cloud  
 $\vec{C}$ : Vector of correlation measurements over time  
 $C_{\max}$ : Maximum correlation value  
 $F_{\max}$ : Maximum font size  
 $F_{\min}$ : Minimum font size

10

Group criterion: Semantic or statistical similarity

Tag shape: Complies (and thus forms) the group's shape

Tag placement: Placed where the group has its highest correlation significance

Group shape: Describes the group's correlation significance over time

15 Color-coding: -

Interaction: Query restriction based on time period and common group criteria  
(similar tag category)

20 Now there will be described an embodiment which makes use of a reference tag selection  
and correlation degree visualization.

Besides focusing on the significance of a tag over time, the formation of groups of tags

and their correlation significance over time, the *degree* of the correlation *between* tags is another valuable meta-information of the represented data. By applying a certain color schema, such correlations can be made visible to indicate simple co-occurrences, semantic similarities or statistical interpretations. To highlight correlations between tags, a reference tag is selected. Assigning a certain reference color to its font with 100% saturation indicates the selection of the tag. All correlated tags are thus highlighted by the same color as the selected tag, whereas the degree of the correlation with the selected tag is equivalent to the saturation of the color applied to the correlated tag. This means that a high degree of correlation is indicated by a high percentage of color saturation, whereas a low percentage of color saturation is equivalent to a low degree of correlation. Figure 9 describes a sample tag cloud with “coffee” as the selected tag, in which correlated tags such as “commuting” or “busy” are indicated by a fading color saturation.

The actual selection of a tag may have two different effects. For one, the described correlation degree visualization (as described above) may focus on correlation dependencies within the given cloud. Another feasible pattern is a correlation-based navigation within the dataset. Here, whenever a certain tag has been selected, a query may be compiled to retrieve only correlated tags. This way, only a subset of correlated information is regarded to provide efficient means to find meaningful information.

20

Now there will be described an embodiment which makes use of temporal correlation degree visualization.

By utilizing inner-tag white-spacing or font-spacing, not only a relative time period can be encoded within a single tag’s representation (as exemplified before), but alternatively its

25

correlation degree over time with (respect to a reference tag) can be described. Here, an increasing inner-tag space indicates a decreasing correlation degree and vice versa. No variation of the inner-tag space indicates a steady correlation degree over time. The average correlation degree of two tags can then also be weighted, whereas recent correlation degrees have a higher weight. This is schematically illustrated in Fig. 10, where the reference tag is “commuting” and the increasing white space of the tag nuremberg indicates that its correlation with the tag “commuting” is decreasing over time. On the other hand, for the tag “working” the correlation with the term “commuting” is increasing over time, as indicated by the decreasing white spaces between the characters.

For each tag further the overall correlation degree (e.g. averaged over time, where the more recent correlation values may be given a higher weight) are indicated, as can be seen by the individual single lines in y-direction. It can be seen that the overall correlation with the tag “commuting” is about the same for the tags “transportation” and “working”, and it is somewhat lower for the tag “nuremberg”.

The white-spacing correlation degree distribution can be described by:

$$\vec{S}(\vec{C}) = [s_0, \dots, s_k] \text{ with}$$

$$\forall_{i=0}^k s_i = S_{\min} + \left(\frac{c_i}{C_{\max}}\right) \times (S_{\max} - S_{\min}),$$

$$(S_{\max} - S_{\min}) > 0,$$

$$\vec{C} = [c_0, \dots, c_k] \text{ with } \forall_{j=0}^k c_j \cdot 0 \leq c_j \leq C_{\max}$$

With:

$s_i$  : Space in between two inner-tag's sections (e.g. letters)

$S_{\min}$  : Minimum (default) space between two inner-tag's sections (e.g. letters)

$S_{\max}$  : Maximum space between two inner-tag's sections (e.g. letters)

$c_j$  : Measured correlation degree for the concerned tag

5  $C_{\max}$  : Maximum correlation degree

The average weighted correlation degree  $A$  can be described by:

$$A(\vec{C}, W) = \sum_{i=1}^k \left( \frac{c_i \times W}{i} \right) \text{with}$$

$$\vec{C} = [c_1, \dots, c_k] \text{with } \forall_{i=1}^k c_i, 0 \leq c_i \leq C_{\max},$$

$$W > 0$$

10 With:

A: Average, weighted correlation degree

$\vec{C}$ : Vector of correlation measurements of a given tag over time

$C_{\max}$ : Maximum correlation value

W: Weighting factor

15

Once a reference tag ("commuting") has been selected, a distinct number of correlated tags is placed above, slightly rotated to intensify the visual effect of increasing or decreasing distance between the reference tag and the correlated tags. A query or browsing based on such a visualization can e.g. be implemented such that whenever the

20 aforementioned reference tag is re-selected, the next group of (less) correlated tags is positioned in a similar manner, replacing the previously shown group of correlated tags.

In the following some of the embodiments which have been explained before will once again be summarily to compiled by illustrating the metainformation which is displayed together with its corresponding graphical visualization.

5

Inner-tag significance distribution	aaaaaaaaa
Inner-tag time period visualization	C o m m u t i n g
In-line grouping based on similar significance distribution of single tags with color-coded correlation significance distribution	veler coffeee munichny
In-line grouping based on similar significance distribution of distinct tags, group's shape forms the common tag criterion	traveler IYOUIT munich coffee
In-line grouping based on semantic or statistical similarity of distinct tags, group's shape forms the group's correlation significance over time	laim munich nuremberg berlin
Color-coding for visualizing the correlation degree between distinct tags and a corresponding reference tag	

Reference tag selection for correlation degree visualization. An increasing inner-tag space indicates a decreasing correlation degree and vice versa. No variation of the inner-tag space indicates a steady correlation degree over time

transportation  
 n u r e m b e r g  
 w o r k i n g  
**commuting**

In the embodiments described so far there have largely been used three ways of visualizing the metainformation for a tag, namely the font size, the color saturation, and the inner space between said slices. It should however be mentioned that also further ways may be used to visualize the metainformation for the individual slices of the tags in order to display its temporal change. Any suitable way for visualizing graphically for a slice of a tag an information which corresponds to a certain quantity of the metainformation associated with the tag may be used. One further such example is illustrated in Fig. 11 where tag's atomic elements (the "slices") are placed on a virtual stairway, whereas one distinct significance measurement is indicated by the actual height with respect to a baseline. Therefore, an element that is more significant is placed higher than less significant elements. The computation of the actual range of height values is similar to the previous font size computation and can be described by:

15 With:

$\vec{L}$ : Vector of distinct sections of a given tag

$\vec{S}$ : Vector of significance measurements of a given tag over time

$S_{\max}$ : Maximum significance value

$H_{\max}$ : Maximum height with respect to baseline

20  $H_{\min}$ : Minimum height with respect to baseline

The formula below maps the given range of significance values to the respective range of height values of a distinct tag, described by the resulting height vector  $\vec{H}$ .

25  $\vec{H}(\vec{L}, \vec{S}) = [h_0, \dots, h_k]$  with

$$\forall_{i=0}^k h_i = H_{\min} + \left( \frac{S_i}{S_{\max}} \right) \times (H_{\max} - H_{\min}),$$

$$(H_{\max} - H_{\min}) > 0,$$

$$\vec{S} = [s_0, \dots, s_k] \text{ with } \forall_{j=0}^k s_j, 0 \leq s_j \leq S_{\max},$$

$$\vec{L} = [l_0, \dots, l_k]$$

Since the time intervals between historic measurements of significance values may also  
 5 be different, the inner-tag space may indicate this temporal divergence. In one  
 embodiment smaller spaces are equivalent to a shorter time interval, larger spaces mean  
 longer time intervals, as illustrated in the upper left part of Fig. 11.

It has been mentioned already that in principle the use of different fonts for a single tag for  
 10 indicating either the significance, correlation distribution or a certain time interval is also  
 possible (this was explained in an earlier embodiment, where as an example is given  
 which mentions the use of the fonts Arial Regular & Arial narrow).

In addition to this, however, also a special type of font could be used, as illustrated in  
 15 Figure 12 for the tag "docomo". In case of such a font the size even inside a character  
 may change, and such a font may be constructed by providing individual slices for each  
 character which then may change their size in accordance with the significance value to  
 be visualized, as can be seen from Fig. 12.

20 More generally speaking one may therefore say that in one embodiment the visualization  
 of the meta-information is performed by changing a visualization parameter for individual  
 the slices of a tag based on the quantity of the meta-information, where the visualization

parameter is chosen such that for the slice of a font it is capable to indicate such a quantity of a metainformation.

It will be understood by the skilled person that the embodiments described hereinbefore  
5 may be implemented by hardware, by software, or by a combination of software and  
hardware. The modules and functions described in connection with embodiments of the  
invention may be as a whole or in part implemented by microprocessors or computers  
which are suitably programmed such as to act in accordance with the methods explained  
in connection with embodiments of the invention. An apparatus implementing an  
10 embodiment of the invention may e.g. comprise computing device or a mobile phone or  
any mobile device which is suitably programmed such that it is able to carry out a  
delegated transaction as described in the embodiments of the invention.

According to an embodiment of the invention there is provided a computer program,  
15 either stored in a data carrier or in some other way embodied by some physical means  
such as a recording medium or a transmission link which when being executed on a  
computer enables the computer to operate in accordance with the embodiments of the  
invention described hereinbefore.

**What is claimed is:**

1. A computer-implemented method of visualizing a tag cloud of individual tags, wherein each is a certain piece of information and is formed of one or more characters, and  
5 wherein for a tag metainformation has been measured over a certain time interval so that there exists a vector of metainformation measurements for said tag, each component of said vector representing a certain metainformation measurement, said method comprising:  
dividing a tag into a plurality of slices;  
10 mapping said vector of metainformation measurements of said tag over time such that each slice of said tag corresponds to one or more of said metainformation measurements; visualizing the temporal change of said metainformation measurements for said tag by changing for each of said slices one or more of the following parameters depending on the corresponding one or more metainformation measurements:  
15       the font size,  
          the color saturation;  
          the inner space between said slices;  
          the actual height on which the individual slices are placed on a virtual stairway, whereas one distinct significance measurement is indicated by the actual height  
20       with respect to a baseline.
2. The method of claim 1, wherein said metainformation is one of the following:  
          the significance of a tag,  
          the correlation significance between tags.
- 25
3. The method of claim 1 or 2, wherein  
          the total period of time which is mapped onto a certain tag is graphically represented by the total length of said tag.
- 30
4. The method of one of the preceding claims, further comprising:  
          identifying for each tag to which temporal pattern of a plurality of predefined temporal pattern it belongs, and  
          grouping the tags together which belong to the same pattern.

5. The method of one of the preceding claims, wherein a plurality of tags which belonging to the same group of tags are selected by  
a similarity criterion, or  
5 arbitrarily by a user, and wherein for the tags of said group metainformation is displayed by one or more of the following:  
a change in font size;  
a change in color saturation.
- 10
6. The method of one of the preceding claims, wherein a plurality of tags which form a group is aligned in a row and said group is divided into slices, each slice representing a certain time period, wherein the vector of metainformation to be visualized is  
representing the correlation significance between said elements of said group over time,  
15 said vector of metainformation being mapped onto said slices of said group of tags such that the visual representation of said group of tags illustrates the change of the correlation between the tags of said group over time based on a change of font size or color saturation.
- 20
7. The method of one of the preceding claims, comprising:  
selecting a group of tags based on a similarity criterion and aligning them in a row and said group being divided into slices, wherein the overall shape of the group is based on the temporal change of the metainformation which corresponds to the similarity  
criterion, and/or  
25 with the tags of the group being ordered such that the tag of the group which has the highest individual significance is located at the position with the highest font size or color saturation and the other tags are ordered according to their correlation with the tag having the highest significance.
- 30
8. The method of one of the preceding claims, wherein  
a group of tags has been selected based on a similarity criterion according to which the members of said group have a certain similarity, and  
the overall correlation significance of the correlation between the group members over

time is visualized by

a changing font size which changes in accordance with the correlation significance change over time; or

5 a color saturation change which changes in accordance with the correlation significance change over time.

9. The method of one of the preceding claims, comprising:

selecting one tag as reference tag;

10 graphically visualizing the correlation of other tags with said selected tag by displaying said other tags such that the degree of correlation with said reference tag is expressed by

the font size of said other tags; or

the color saturation of said other tags.

15 10. The method of claim 9, wherein

the inner significance of a tag over time is expressed by one of its font size or color saturation; and

the correlation significance of a tag with said selected tag over time is represented by the other one of said font size or said color saturation.

20

11. The method of one of the preceding claims, wherein a reference tag is selected and other tags are simultaneously displayed in a manner which indicates their respective correlation with said selected tag over time, wherein said correlation is indicated by the space between slices or characters of a tag are changing in accordance with the

25 change of said correlation, or

the color saturation changing in accordance with the change of said correlation.

12. The method of claim 11, further comprising:

30 in addition to visualizing the correlation of a tag with said reference tag over time, further visualizing the overall correlation of said tag with said reference tag.

13. The method of one of the preceding claims, wherein

said visualization of said tag cloud enables the user to query or browse context

information as represented by said tag cloud based on the metainformation represented in said tag cloud, said metainformation being one or more of the following:

temporal change of a tag significance over time,

temporal change of a correlation significance between tags over time,

5 any metainformation about the tags of said tag cloud as graphically visualized by said tag could.

14. A computing apparatus for visualizing a tag cloud of individual tags, wherein each is a certain piece of information and is formed of one or more characters, and wherein for a  
10 tag metainformation has been measured over a certain time interval so that there exists a vector of metainformation measurements for said tag, each component of said vector representing a certain metainformation measurement, said apparatus comprising:

means for dividing a tag into a plurality of slices;

15 means for mapping said vector of metainformation measurements of said tag over time such that each slice of said tag corresponds to one or more of said metainformation measurements;

means for visualizing the temporal change of said metainformation measurements for said tag by changing for each of said slices one or more visualization parameters, wherein said visualization parameters preferably comprise one or more of the following  
20 parameters depending on the corresponding one or more metainformation measurements:

the font size,

the color saturation;

the inner space between said slices;

25 the actual height on which the individual slices are placed on a virtual stairway, wherein one distinct significance measurement is indicated by the actual height with respect to a baseline.

15. A computer program comprising computer program code which when being executed  
30 on a computer enables said computer to carry out a method according to one of claims 1 to 13.

## **1. ABSTRACT**

A method of visualizing a tag cloud of individual tags, wherein each is a certain piece of  
5 information and is formed of one or more characters, and wherein for a tag  
metainformation has been measured over a certain time interval so that there exists a  
vector of metainformation measurements for said tag, each component of said vector  
representing a certain metainformation measurement, said method comprising:  
dividing a tag into a plurality of slices;  
10 mapping said vector of metainformation measurements of said tag over time such that  
each slice of said tag corresponds to one or more of said metainformation measurements;  
visualizing the temporal change of said metainformation measurements for said tag by  
changing for each of said slices one or more of the following parameters depending on  
the corresponding one or more metainformation measurements:  
15       the font size,  
          the color saturation;  
          the inner space between said slices;  
          the actual height on which the individual slices are placed on a virtual stairway,  
          whereas one distinct significance measurement is indicated by the actual height  
20       with respect to a baseline.

**africa** amsterdam animals april architecture art asia australia baby band barcelona **beach**  
 berlin bird **birthday** black blackandwhite blue boston bw **california** cameraphone camping  
 canada canon car cat chicago china christmas church city clouds color concert  
 cute dance day de dog england europe fall **family** festival film florida **flower**  
**flowers** food france friends fun garden geotagged germany girl girls graffiti  
 green halloween hawaii hiking holiday home honeymoon house india ireland island **italy**  
**japan** july kids la lake landscape light live london macro me mexico mountain  
 mountains museum **music** nature new newyork newyorkcity newzealand **night** nikon  
 nyc ocean **paris** park party people photo photos portrait red river rock rome san  
 sanfrancisco scotland sea seattle show sky snow spain spring street **summer**  
 sun sunset sydney taiwan texas thailand tokyo toronto **travel** tree trees trip uk urban  
**usa** vacation vancouver washington water **wedding** white winter yellow york  
 zoo

**Fig. 1**

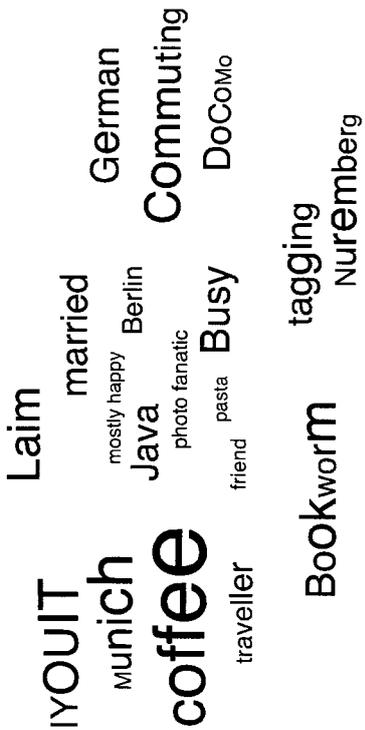


Fig. 4

...	...	...
IYOUIT munich coffee traveller	Laim married mostly happy Java photo fanatic friend	German Commuting DoCoMo
Bookworm	...	tagging Nuremberg

Fig. 5

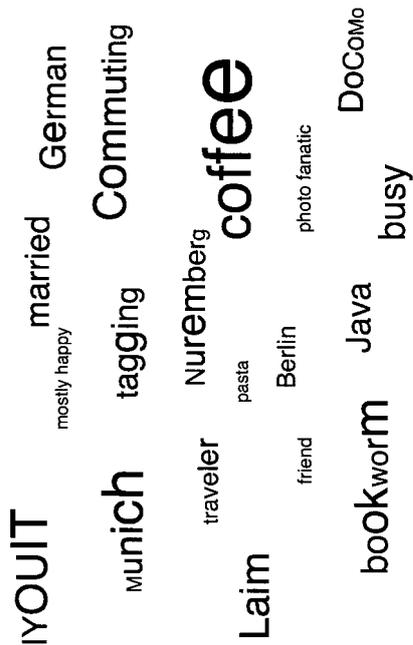


Fig. 2

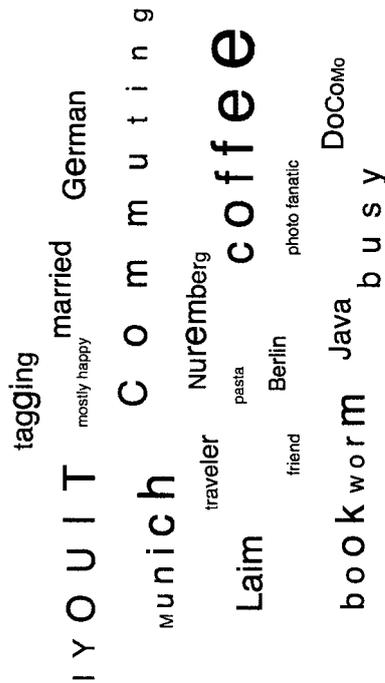


Fig. 3

traveler **coffee** munich **iyouit**

Fig. 6

traveler **IYOUIT** munich **coffee**

Fig. 7

laim munich nuremberg berlin



**Fig. 9**

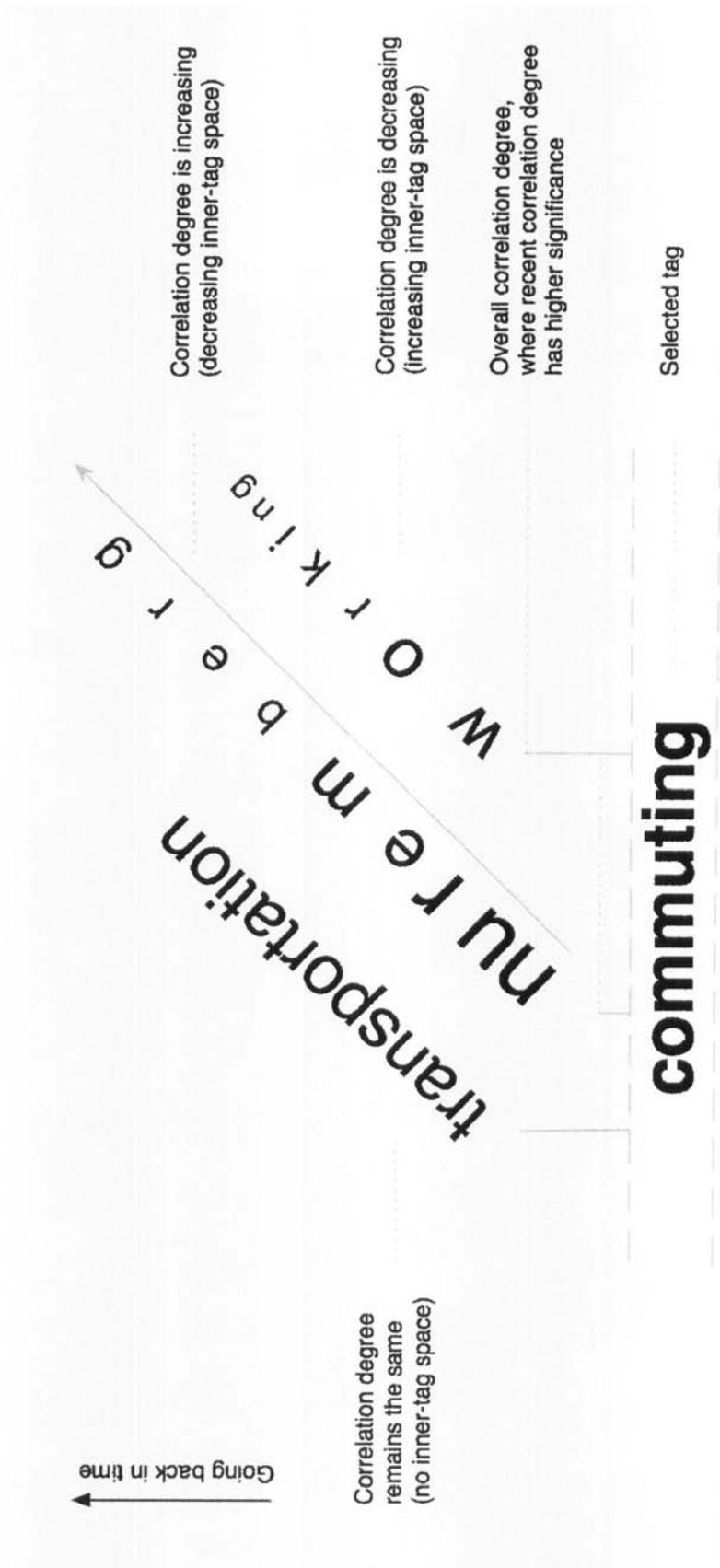


Fig. 10

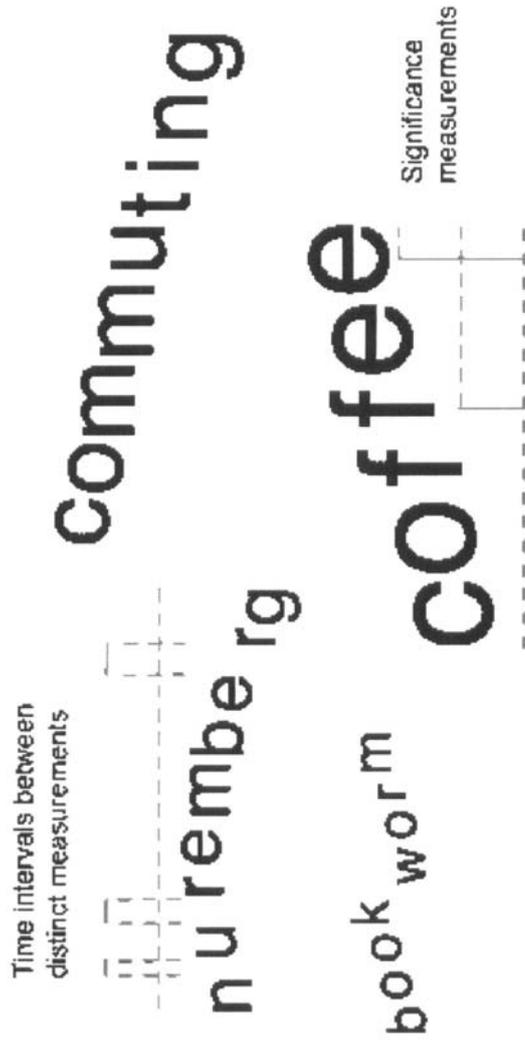
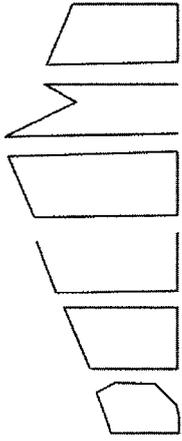


Fig. 11



**Fig. 12**