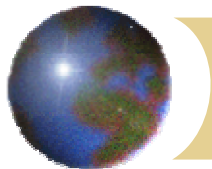


# モバイル環境における 地理情報の関連を利用した Web キャッシュ管理

五島一将<sup>†</sup> 李龍<sup>†</sup> 高倉弘喜<sup>‡</sup> 上林弥彦<sup>†</sup>

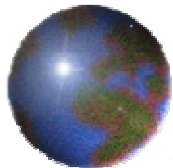
<sup>†</sup>京都大学大学院情報学研究科

<sup>‡</sup>京都大学大型計算機センター



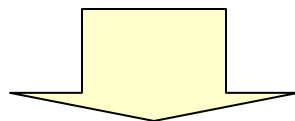
# 発表内容

- ⊕ 研究背景
- ⊕ アプリケーションの提案
- ⊕ KyotoSEARCHの紹介
- ⊕ ページ重要度計算アルゴリズム
- ⊕ シミュレーション・考察
- ⊕ おわりに

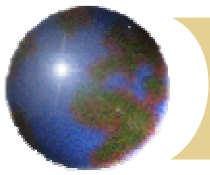


# 研究背景

- ❁ **モバイルWeb検索での制約**
  - ❁ 記憶容量の問題
  - ❁ 通信コストの問題
  - ❁ その他、処理能力・電池の問題

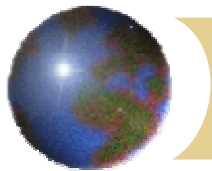


**モバイル端末にWebページを効率よくキャッシュすることの重要性**



# Webページのキャッシュ

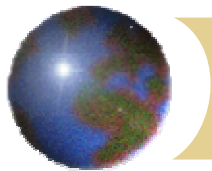
- ❁ 従来のキャッシュ(置き換え)アルゴリズム
  - ❑ LRU, LFU, ...
  - ❑ メモリデータが対象
    - アクセス時間や頻度・・・データの内容は考慮せず
- ❁ Webページキャッシュでは
  - ❑ データサイズ大きい、まちまち
  - ❑ それ自体が「意味」を持つ



# モバイル環境でのWeb キャッシュ

- ✦ ページの持つ「位置情報」が重要
  - ✦ 自分からの距離など、ユーザ位置との関連
- ✦ その他、さまざまなユーザ状況が影響
  - ✦ ユーザの目的・興味
    - ・ 必要とする情報はユーザによって異なる
  - ✦ 行動履歴
    - ・ 一度訪れた場所の情報は不要に
  - ✦ 外部環境
    - ・ 時刻, 天候 etc.

Web ページキャッシュ・・・ユーザにとって重要度の高い / 低いページは？



# 本研究の目的

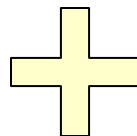
## ✪ モバイル環境に適したWebページキャッシュ

### 📍 旅行中のWeb検索を支援するモバイルアプリケーションの提案

- ページの重要度に応じたレベルのメタデータキャッシュ

従来のキャッシュ

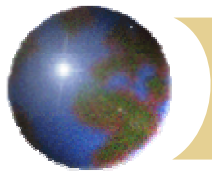
コンテンツそのもの  
テキスト情報のみ



メタデータ

インデックス情報  
URL

- **メタデータを使って**
  - 求めるページを検索 ページをダウンロード
  - 同じような内容のページの重複を避ける



# 本研究の目的

## ✦ モバイルユーザの状態に基づいたWebページ キャッシュ

### ■ ページ重要度順位を求めるアルゴリズム

- 「Webページの持つ位置情報」と「ユーザの現在位置」をもとに判定



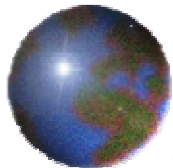
- **地名どうしの関連**を利用

– 地域Web情報検索システム **KyotoSEARCH**

- 種々のユーザ状態を考慮したモデル

### ■ ページ重要度を調べることによって

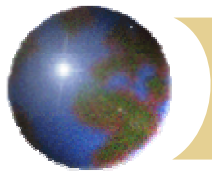
- キャッシュ置き換え・段階的キャッシュ
- 能動的観光ガイド      高重要度ページの推薦・先読み



# 発表内容

- ❖ 研究背景
- ❖ アプリケーションの提案
- ❖ KyotoSEARCHの紹介
- ❖ ページ重要度計算アルゴリズム
- ❖ シミュレーション・考察
- ❖ おわりに





# モバイル環境でのWeb情報の携帯

## 旅行前のプランニング



- モバイル端末の**記憶容量**は限られているため全ての情報を持っていくのは不可能、また必ずしも必要になるとは限らない

## 旅行中の検索・ガイド



- **限られた通信環境**で多くの情報を検索するのは大変

Web Cacheの効果的な利用

必要最小限の情報をモバイル端末に格納して携帯し、屋外で必要に応じて検索

# モバイルに格納する情報

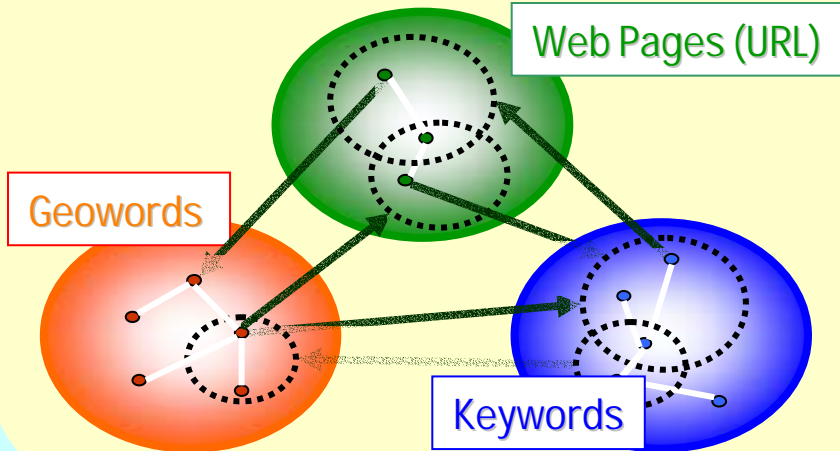
## メタデータ

訪問予定の場所  
geoword

興味あるkeyword  
(旅行の目的)

ユーザ状態  
パラメータ

Webページと地名・キーワード間の関連



Webコンテンツ  
キャッシュ

Cached  
Web Pages

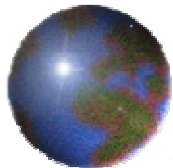
1. Web情報を検索
2. 訪問地・訪問順序の計画を決定
3. メタデータ(URL・関連・その他パラメータ)をモバイル端末に取り込む
4. 必要なWebページをキャッシュに保存

1. モバイル環境でのWeb検索支援 – 現在位置などのユーザ情報を用いた効率的なキャッシュ管理
2. ページ重要度順位に基づく能動的観光ガイド – 付近の関連情報の提供、次訪問地の提案

Planning

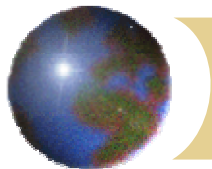


Retrieval &  
Guide



# 発表内容

- ❖ 研究背景
- ❖ アプリケーションの提案
- ❖ KyotoSEARCHの紹介
- ❖ ページ重要度計算アルゴリズム
- ❖ シミュレーション・考察
- ❖ おわりに



# KyotoSEARCHとは

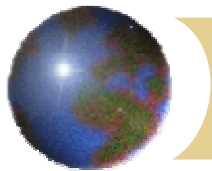
## 地域Web情報検索システム“KyotoSEARCH”

### 特定の地域(京都市)に関するWeb情報の検索・分析を支援

- 京都に関するページの収集
  - 現在約200万ページ
- 名詞を抽出 地名(**geoword**)と非地名(**keyword**)に分類
- それらの関連を分析
  - (現在)ページ内の共起数

### KyotoSEARCH Demo System



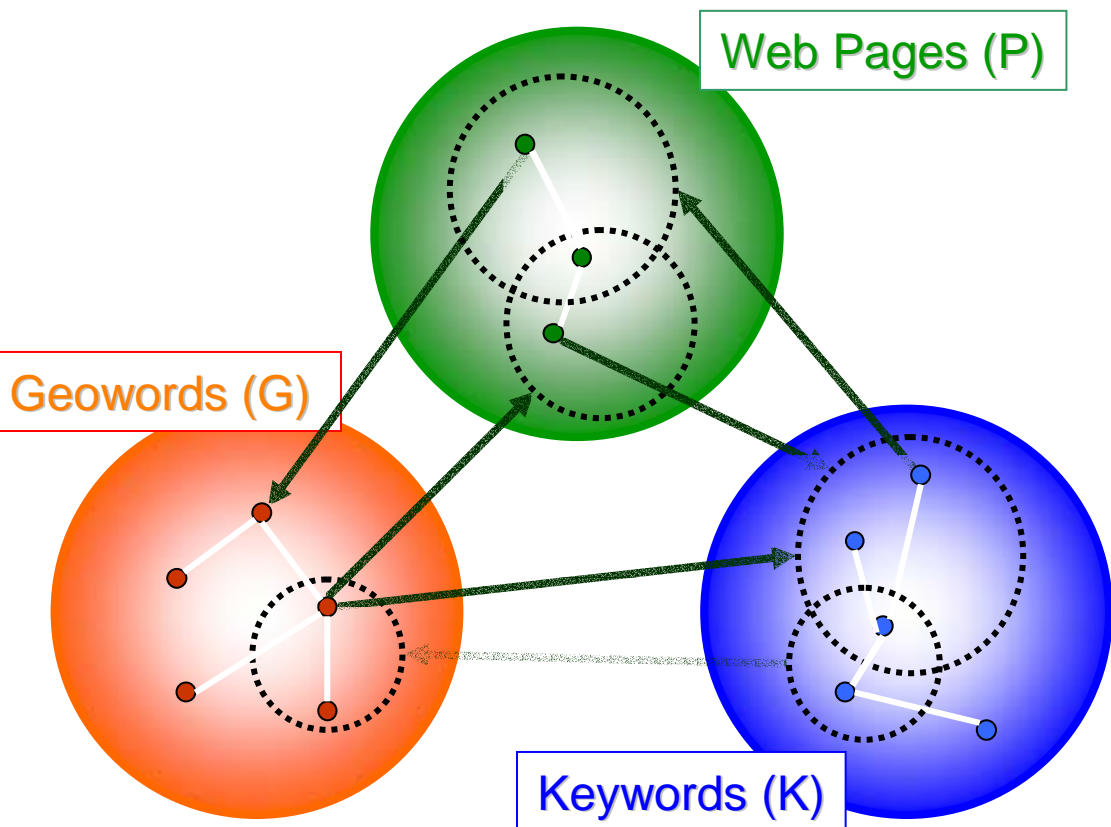


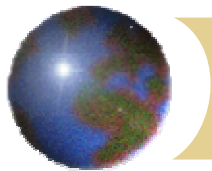
# KyotoSEARCHの基本概念

- KyotoSEARCHが扱うWeb上の関連モデル

Web空間を3つの集合で考える

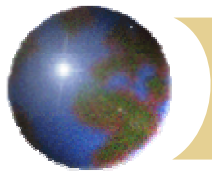
- G, K, P間相互の関連
  - G-G関連, K-K関連
  - G-K (K-G) 関連
  - P-G関連, P-K関連





# 発表内容

- ❖ 研究背景
- ❖ アプリケーションの提案
- ❖ KyotoSEARCHの紹介
- ❖ ページ重要度計算アルゴリズム
- ❖ シミュレーション・考察
- ❖ おわりに



# Web ページ重要度計算

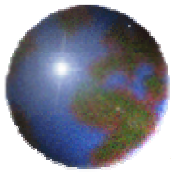
## ◆ 基本モデル:

- ◆ geowordの関連のみを用いたG-Gモデル
- ◆ keywordも用いることのできるモデルに拡張

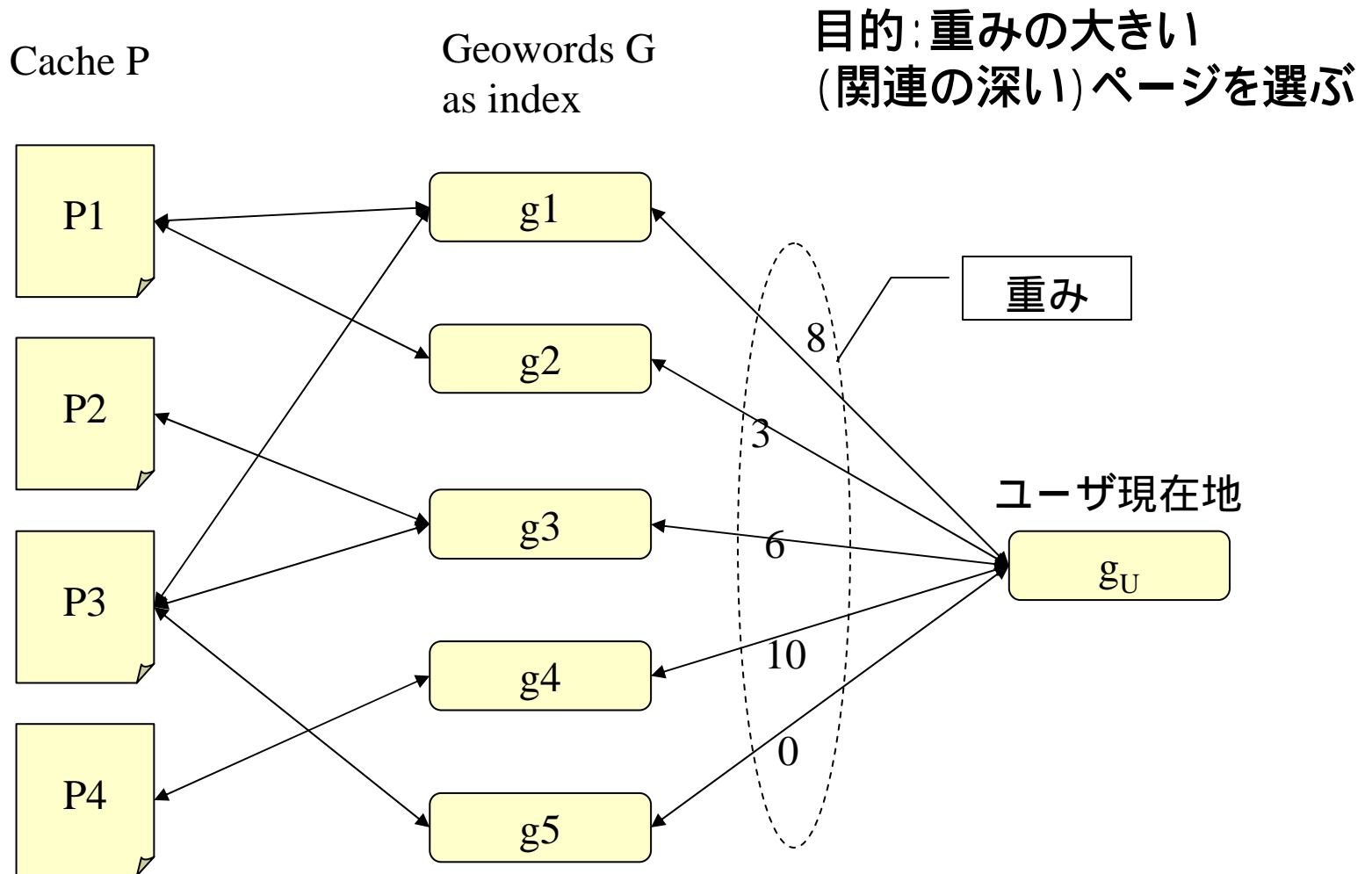
## ◆ その他のユーザ状態の導入

- ◆ 行動履歴や時刻など

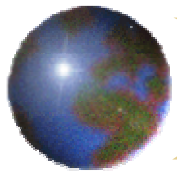




# G-Gモデルによるアルゴリズム

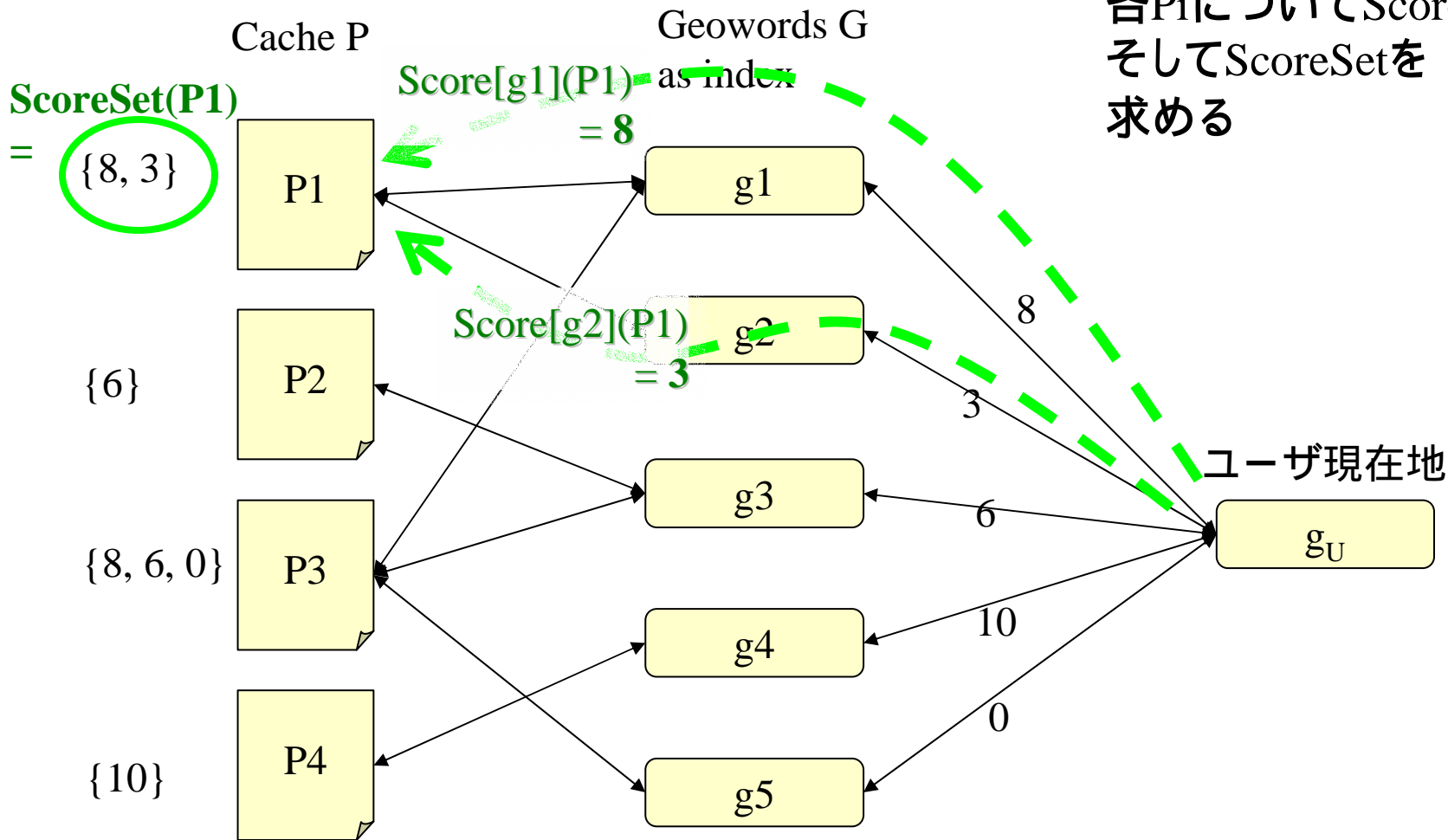


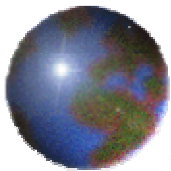




# G-Gモデルによるアルゴリズム

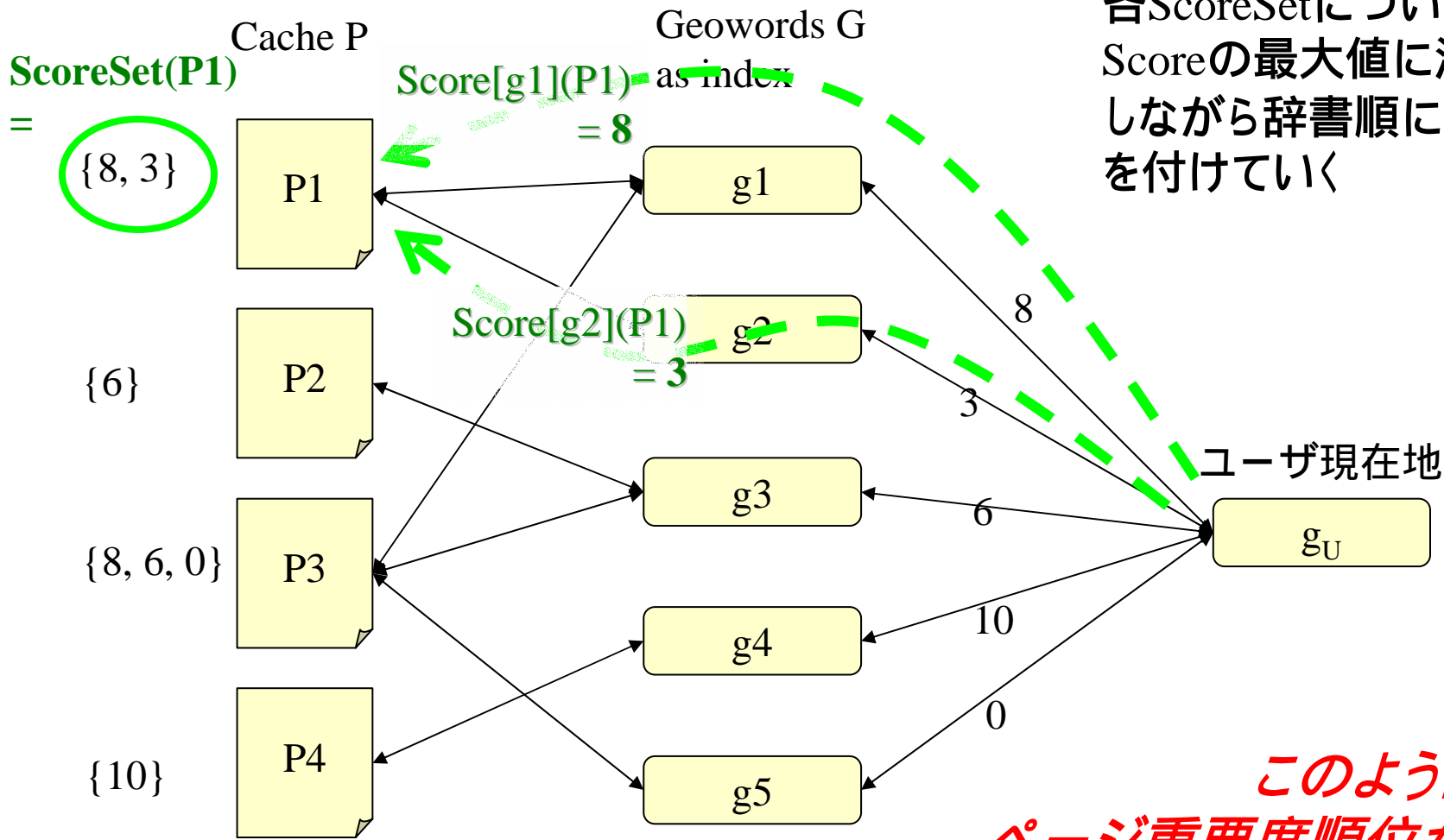
各PiについてScore,  
そしてScoreSetを  
求める



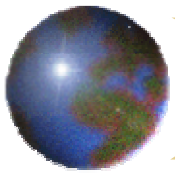


# G-Gモデルによるアルゴリズム

各ScoreSetについて  
Scoreの最大値に注目  
しながら辞書順に順位  
を付けていく

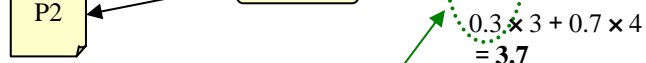
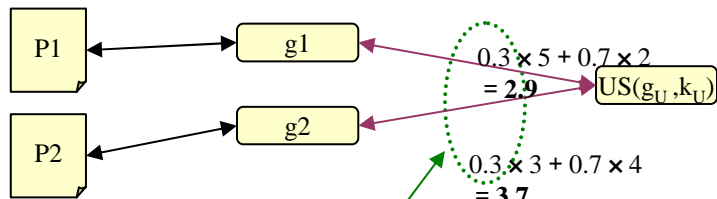
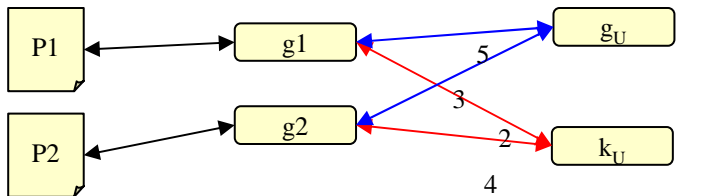


このようにして  
ページ重要度順位が決定

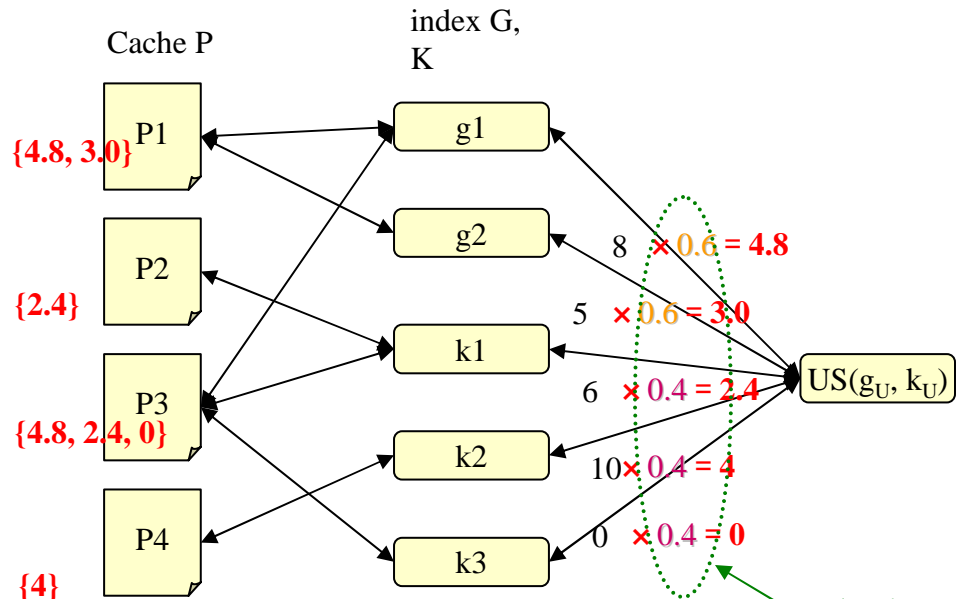


# ユーザ状態: keyword

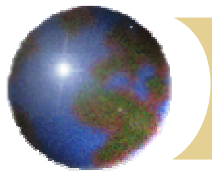
- Keyword: ユーザの興味や目的を示す
  - 計算モデルとしてはgeowordと同等
  - Geowordとモデル統合する際に重視度を設定できる。



重視度



重視度



# ユーザ状態パラメータ

## ✦ Geoword, keywordの関連以外のユーザ状態

### ❑ 行動履歴・訪問順序

- 既に訪れた場所の重要度は低く、これから訪れる場所の重要度は高い

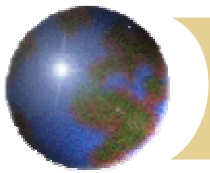
### ❑ 時刻・所持金 etc.

- ユーザが訪れたいと感じる場所が変化する

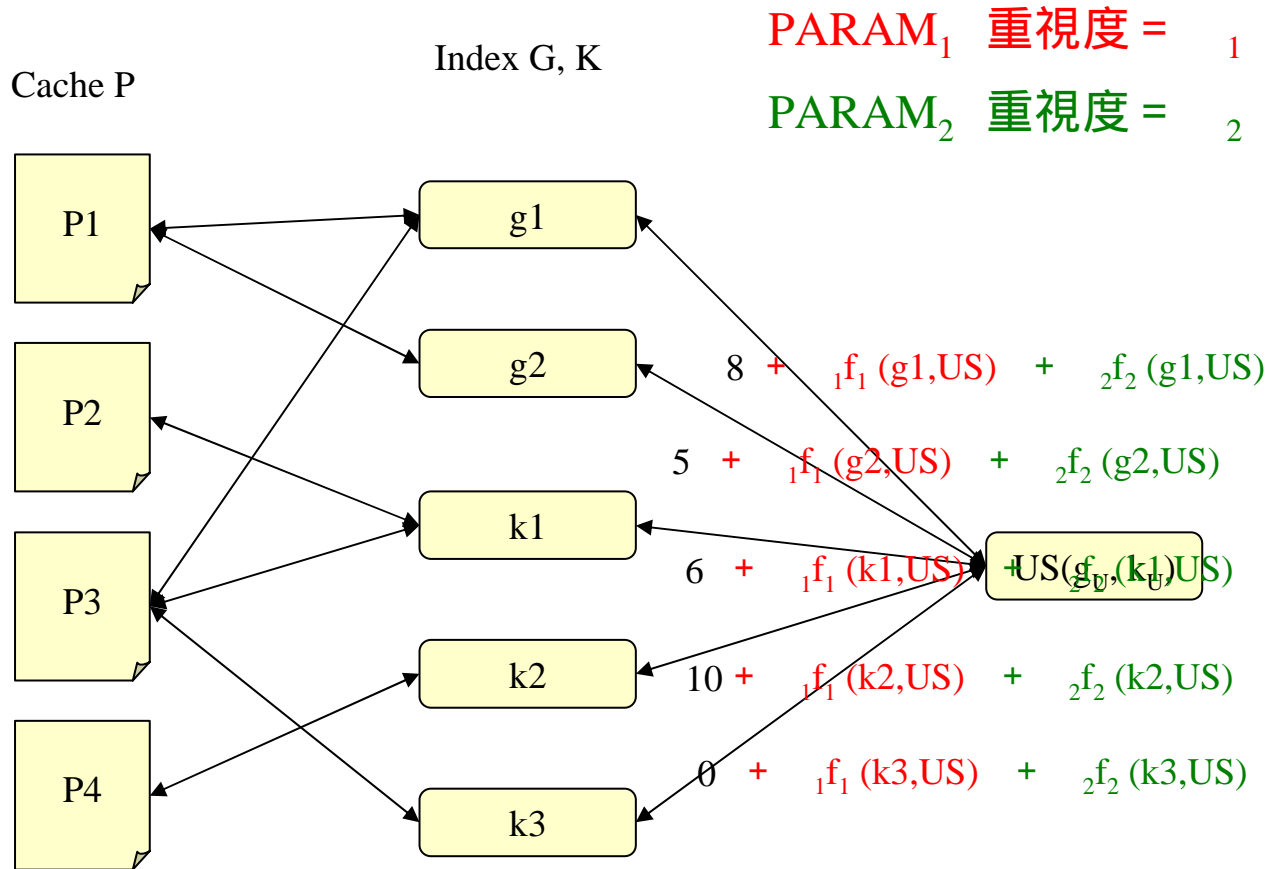
## ✦ パラメータを重要度に反映する方法

### ❑ インデックス (geoword, keyword) とユーザ状態値から決まる関数の値 (増分) をスコアに加算

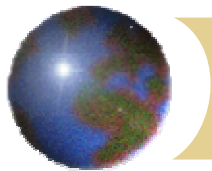
### ❑ 等しい順位となったページにも重要度の差を付けることができる



# パラメータによるスコア増減



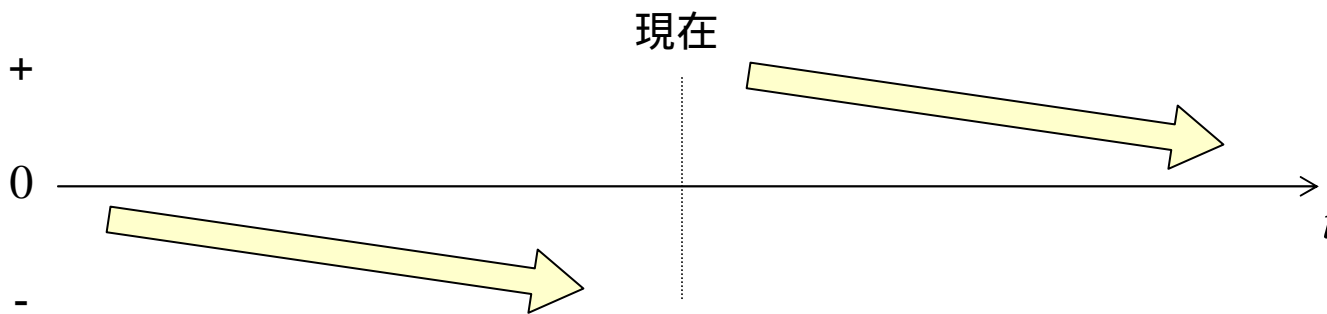
の大きさによりそのパラメータの影響の大きさを調節



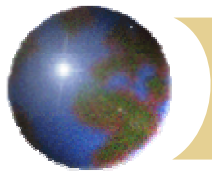
# パラメータの例

## 訪問履歴・訪問順序

- 過去: 最近訪れた場所ほど、再び訪れる必要性がない…重要度低い
- 将来: 訪問予定の近い場所ほど、そこに関する情報の重要性が高い
- 重要度(スコア)の増分は図のように変化:



- $f(\text{geoword}, \text{ユーザ訪問履歴})$  例: 分数関数  $f = \frac{1}{t - t_0}$



# パラメータの例

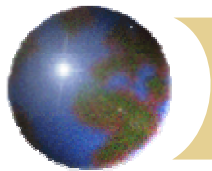
## ❖ 時間帯による重要度変化

- ❑ 飲食店の多い繁華街・・・食事時に需要増
- ❑ 博物館などの施設・・・営業時間
- ❑ 祭りなどのイベント

## ❖ 所持金による重要度変化

- ❑ 高級な店の多い地域      安価・無料の場所
- ❑ 携帯電話による決済・・・所持金管理が可能に
  - ・ 行動履歴からの推測も

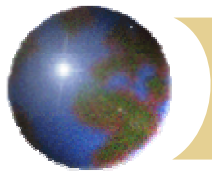
*別DBにより地名と値の対応関数を定義*



# 発表内容

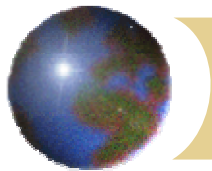
- ✦ 研究背景
- ✦ アプリケーションの提案
- ✦ KyotoSEARCHの紹介
- ✦ ページ重要度計算アルゴリズム
- ✦ シミュレーション・考察
- ✦ おわりに





# 重要度計算のシミュレーション

- ✦ Webページ集合(仮想キャッシュ)に対する順位付け
  - KyotoSEARCHの蓄積ページの中から無作為に選んだ100, 500, 1000ページを対象
  - G-Gモデル(地名どうしの関連)で計算
  - ユーザ現在地:「銀閣寺」と「祇園」の2通りを仮定



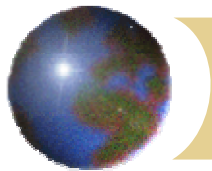
# 結果と考察

## ✦ 上位に現れた主なURL

- [www.momonga.org/kyoto/spot-j.html](http://www.momonga.org/kyoto/spot-j.html)
- [gourmet.yahoo.co.jp/gourmet/restaurant/Kinki/Kyoto/list/area\\_genre/260005\\_0207](http://gourmet.yahoo.co.jp/gourmet/restaurant/Kinki/Kyoto/list/area_genre/260005_0207)
- [www.kyoto-np.co.jp/kp/topics/2000jun/bk\\_index.html](http://www.kyoto-np.co.jp/kp/topics/2000jun/bk_index.html)

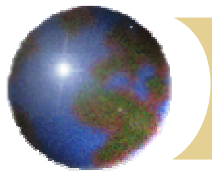
❑ 多くの地名を持つ一覧表的ページが多い

❑ ページの内容に深く踏み込んだ解析が必要



# まとめ

- **モバイル環境でのWebページキャッシュ手法**
  - **モバイル旅行ガイド支援システム**
    - メタデータとWebページキャッシュの携帯
    - ページ重要度を用いた能動的観光ガイド機能
  - **位置情報の関連を用いたWebページの重要度判定**
    - KyotoSEARCHによりWebから抽出、解析した地名やキーワードの関連情報を利用
    - その他のユーザ状態パラメータによるスコア操作



# 今後の課題

- さらなる評価、実装
- ページの「質」によるフィルタリング
  - 多くの地名を含むページがどうしても有利
  - 扱うエリアの網羅度(広さ)、内容の詳細度(深さ)など