

GOOGLEを支える技術

名城大学工学部情報工学科

小田耕平

書籍紹介

- Googleを支える技術
ー巨大システムの内側の世界
技術評論社
西田圭介 著



Googleとは

- Web**検索エンジン**
- 1998年に米国の二人の学生によって作られた。
- いか**に役立つ情報を見つけられるかを重視。**

役立つ情報を見つけるために(1)

□ Page Rank

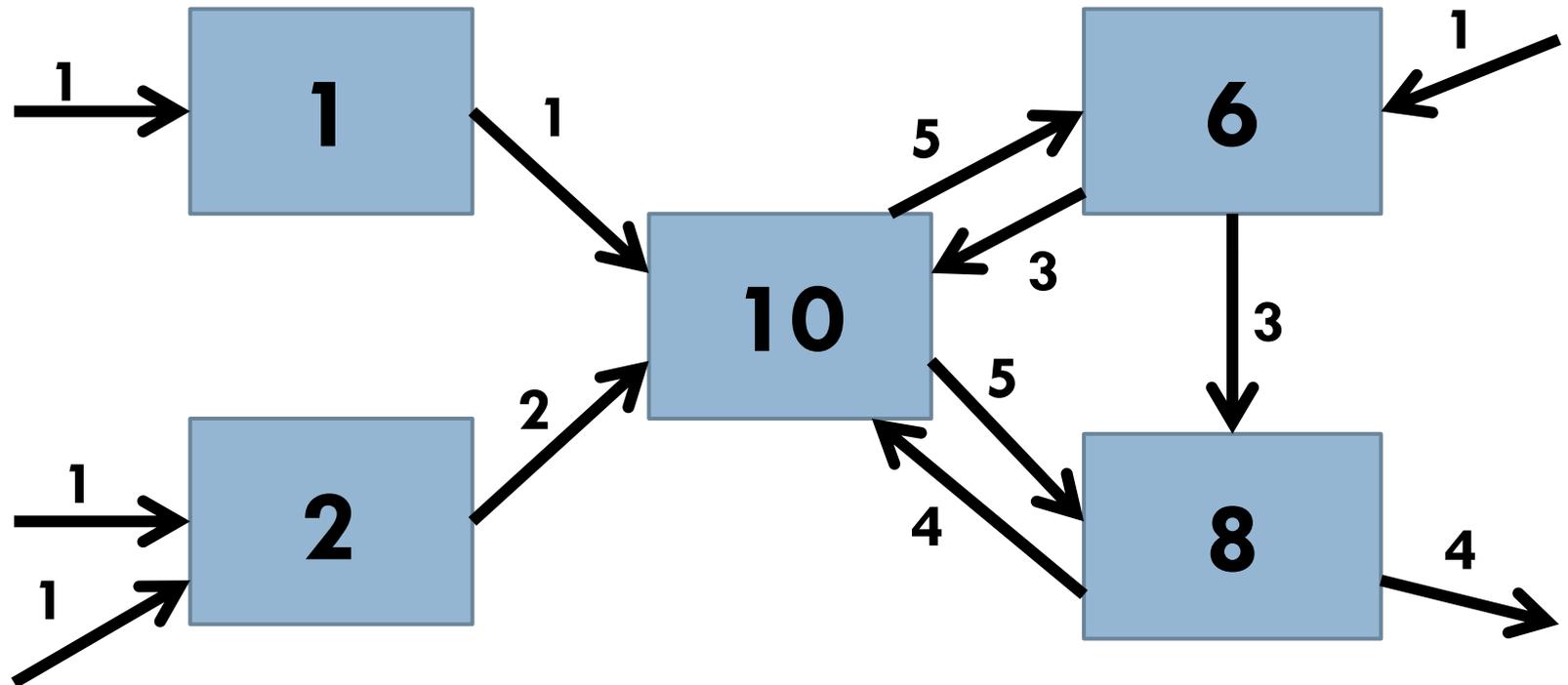
役にたつページはあちこちからリンクされている

□ 概念

- ◆ 各webページは自分の点数を持つ
- ◆ 他のページにリンクすると、自分の点数を配分する。
- ◆ 自分の点数は、他のページからもらった点数の合計で決まる。

役立つ情報を見つけるために(2)

□ 概念図



役立つ情報を見つけるために(3)

□ アンカーテキスト

リンクに付けられた文字列

ユーザーがそれを検索しようとした時にだけ意味を持つ

□ 例

名城大学はいい大学だね。

役立つ情報を見つけるために(4)

- **単語(単語情報)による検索**

webページに含まれるすべての単語を記録
複数の単語で検索

- **例**

“京都 大学”

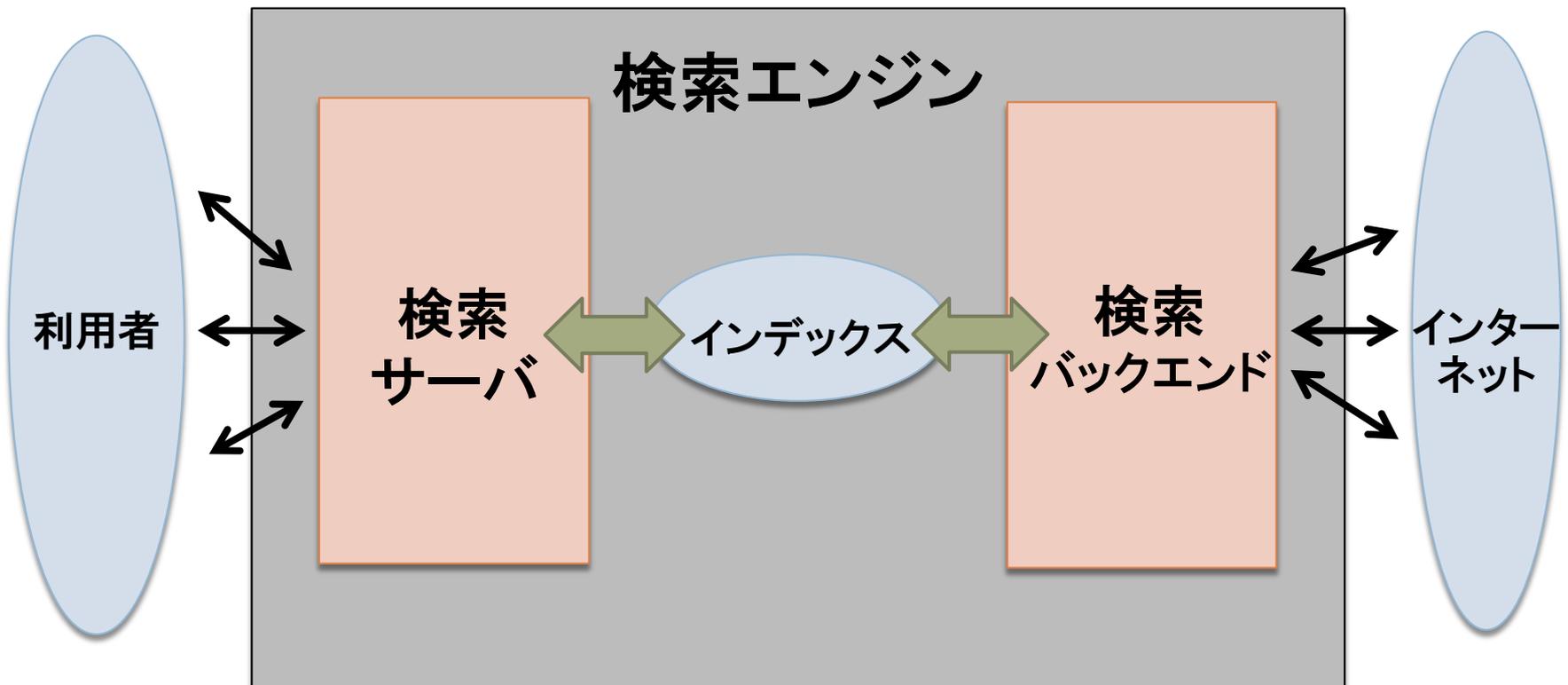
“大学 京都”

前半まとめ

- 役立つ情報を見つけるためには
 - ◆ Page Rank
 - ◆ アンカーテキスト
 - ◆ 単語による検索

検索エンジンのしくみ(1)

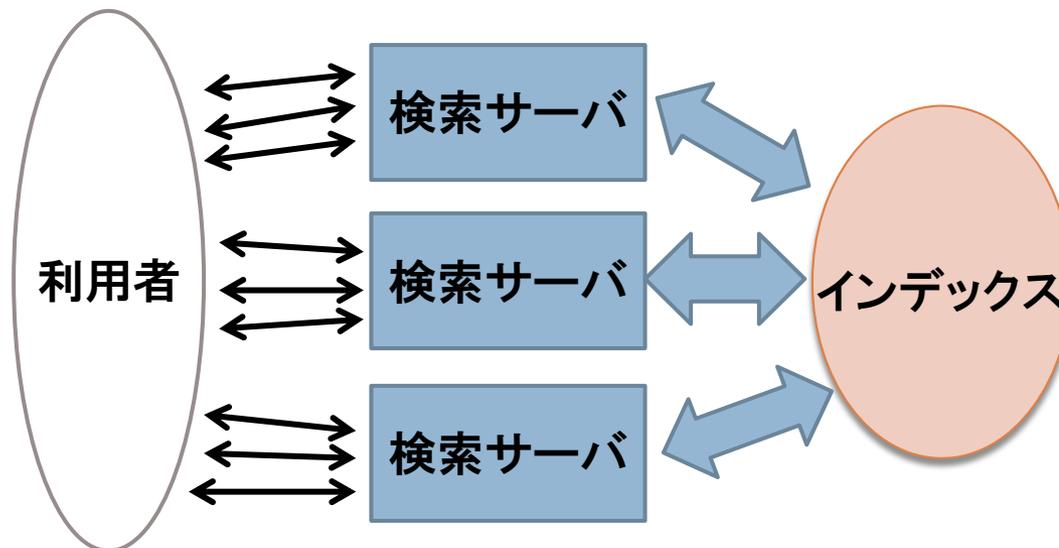
□ 検索エンジンの基本構造



検索エンジンのしくみ(2)

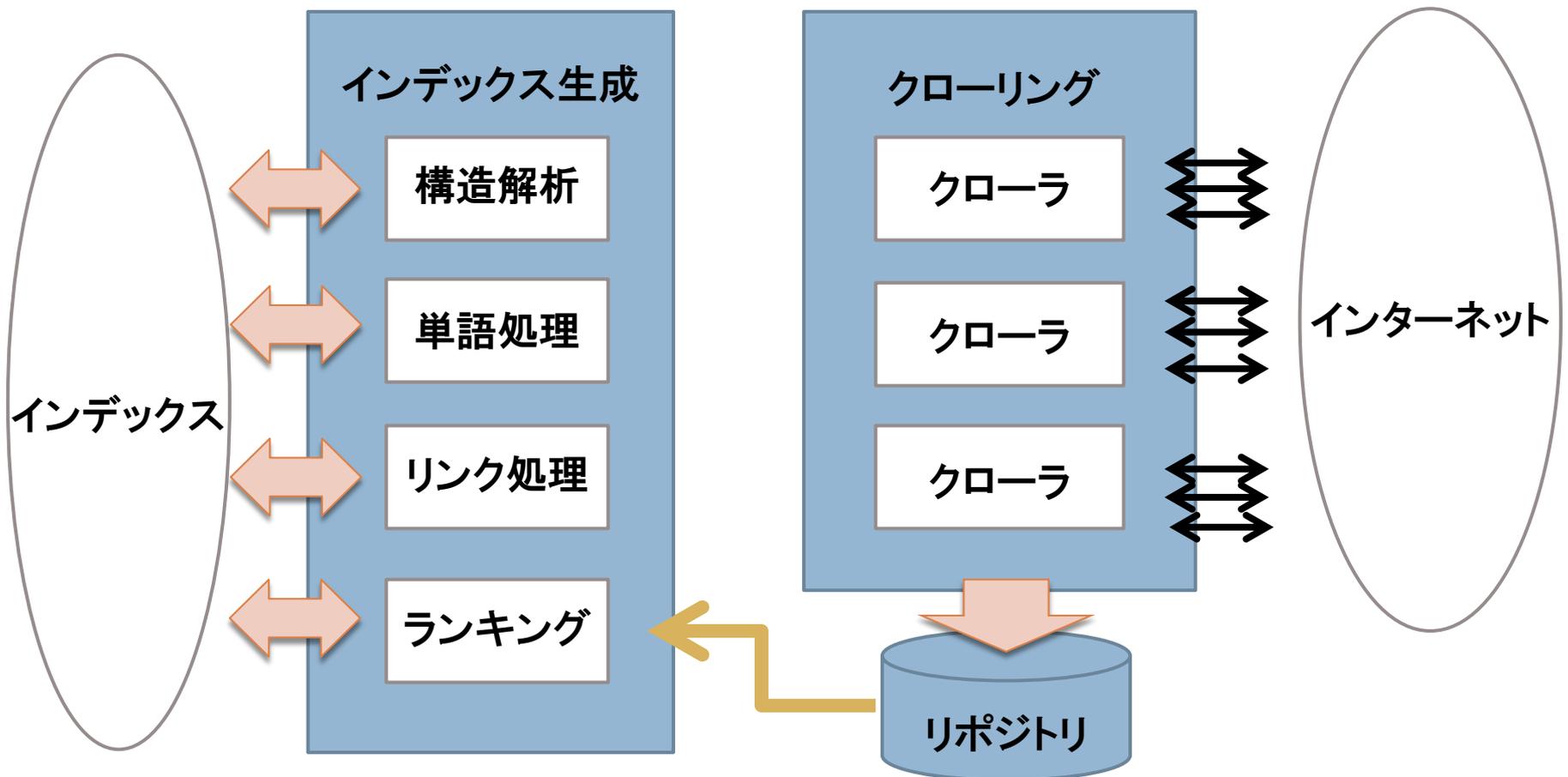
□ 検索サーバ

- ◆ 利用者との通信を管理する
- ◆ 利用者からのリクエストを解析し、行うべき処理を判断する。
- ◆ インデックスから必要な情報を管理する。
- ◆ 結果を見やすくレイアウトし、利用者へ送り出す。



検索エンジンのしくみ(3)

□ 検索バックエンド

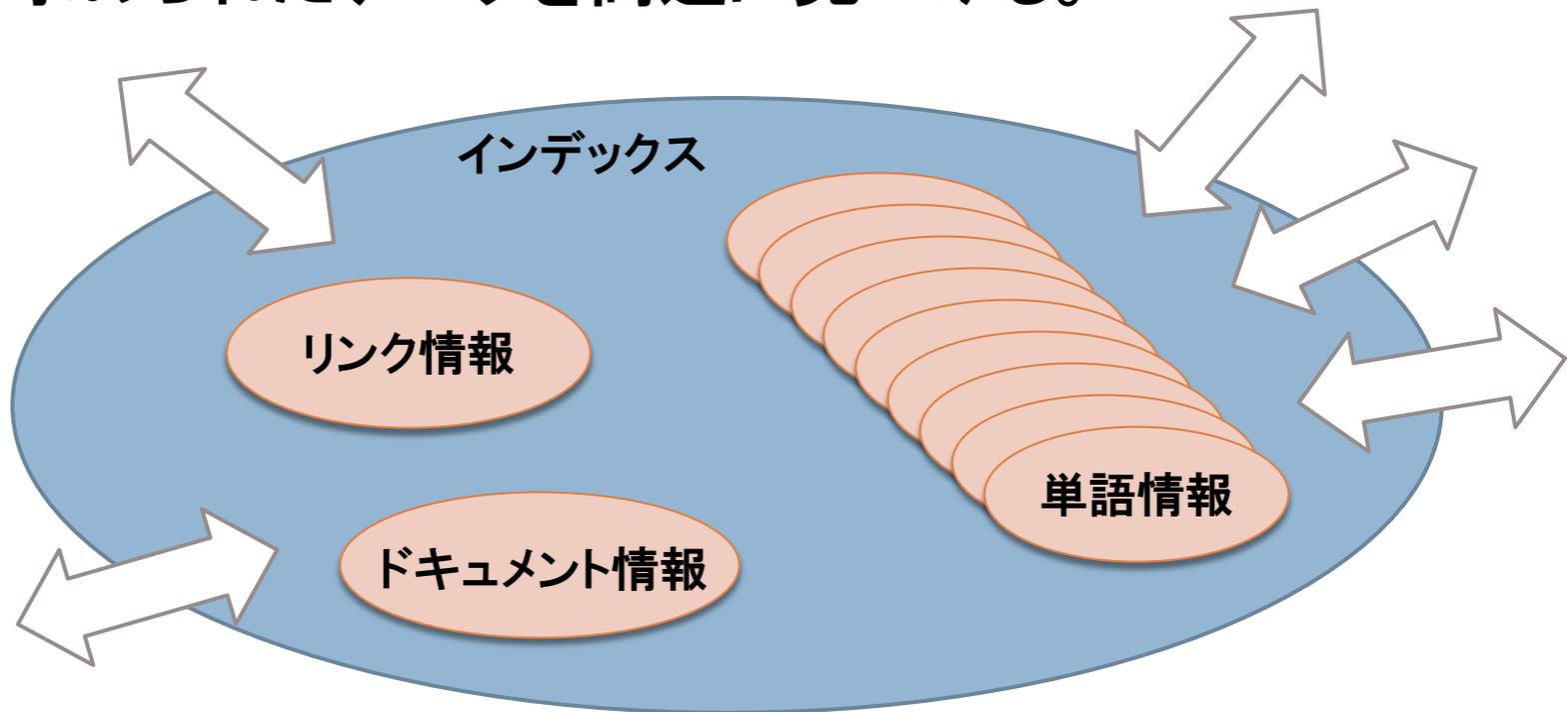


検索エンジンのしくみ(5)

□ インデックス

データを安全に格納する。

求められたデータを高速に見つける。



検索エンジンのしくみ(6)

□ インデックスの構造

検索語	大学
タイトル	名城大学
URL	http://meijo/



キー	値1
大学	101

キー	値
11	名城大学
21	http://meijo/



キー	値1	値2
101	11	21

効率化をはかるポイント

- あらかじめ検索バックエンドによってwebページの情報を解析しておく。
- インデックスに格納する際に数値を使って格納を行う。
- 3つの構造の連携

まとめ

- 検索エンジンのしくみ
- 検索サーバ
- 検索バックエンド
- インデックス



ご清聴ありがとうございました。