WebCluster - Motivation

Information Need

Query -> Infoseek

Gulfs
- information need <-> query
- WWW (search engine) <-> WWW (subject domain)
- unstructured WWW <-> structured subject domain (non-WWW)

WebCluster - Mediated Access

Structured Source Collection

Query -> Infoseek

Source Collection (domain)
Clustering a document collection: an example

Fixed Plants
- Coastal Wind Farms
  - Pacific Rim Wind Farms
  - Design of Coastal Wind Farms
- Inland Wind Farms
  - Desert Wind Farms
- Portable Generators
  - Wind generators for yachts

Wind Energy
- Power Generation
- Propulsion

Mediating access to the WEB: a Concrete Example

1. Clustering
2. Cluster-based search + browsing
3. Query
4. Result
5. Specialised Document Collection (e.g. Wind Power)
WebCluster - Applications

- Structured portals on the WWW e.g. wind energy portal
- CD-ROM based collections
  - structured access to the collection itself
  - mediated access to WWW (via CD-ROM)
  - added-value
- mediated access via hierarchically structured information sources. Examples are: via large structured report (e.g. government reports), via structured collection of information (e.g. encyclopaedia)
- multimedia information access
  - cluster multimedia source, e.g. annotated photographs
  - mediated access to other photographs (not annotated)

Expected Outcomes

- Enable information providers to produce high-quality, specialised, classified collections of selected WWW pages.
- Give end-users effective access to these clustered collections using state-of-the-art data visualisation techniques and user interfaces purpose-designed for information seeking tasks.
- Extend our understanding of cluster-based searching in a networked environment.
- Produce software prototypes which have the potential to be exploited commercially.