access of rich, semantic information by humans or machines

re-publish

data to knowledge

extract, aggregate, link

unstructured and semi-structured distributed data
- Discover and extract Named Entities quickly from the Web
- Assess entity discoveries and maintain high quality knowledge base
- Enhance information about entities with
  - Facts
  - Question/Answers
  - Snippets
  - Events
  - Multimedia objects
WebKnox Architecture
Entity Extraction

Knowledge Ontology

Phrase Extraction

Focused Crawl Extraction

Seed Extraction

Search Engine

Knowledgebase

Assessment

Concept Names

"list of CONCEPT"
"CONCEPT list"
"index of CONCEPT"
"CONCEPT index"
"browse CONCEPT"

"ENTITY1" "ENTITY2"

"CONCEPT such as"
"CONCEPT like"
"CONCEPT including"

Entities

Seed Entities
Motorola RAZR v3i

- Camera: 1.23 mega pixel
- 3G: no
- WLAN: no
- Bluetooth: yes
- GPRS: 10
- USB: yes