1. Introduction

Majority of information retrieval and web search use the inverted index as the backbone for their keyword based search. Unfortunately, inverted index fail to capture the semantic of the language and thus keyword based approach can only go as far as giving the relevant hits. I propose semantic indexing scheme which can encode natural languages without losing semantic relationships.

The encoding of natural language would involve parsing of each sentence, deriving its semantic model and storing the encoded representation. In the below section I discuss the appropriate schema to store the derived semantic model, such that semantic information can be recovered accurately.

2. Schema

Events
id

Actions
id
action

Objects
id
object

Adjectives
id
adjective

Adverbs
id
adverb

Tenses
id
tense

Event-Actions
id
event_id
action_id
tense_id

Event-Agents
id
event_id
agent_id(object_id)

Event-Patients
id
event_id
patient_id(object_id)

Event-Instruments
id
event_id
instrument_id(object_id)

Event-Beneficiaries
id
event_id
beneficiary_id(object_id)

Semantic Indexing

Anant Bhardwaj ANANTB@CSAIL.MIT.EDU

Event-Locations
id
event_id
location_id(object_id)

Event-Adverbs
id
event_id
action_id
adverb_id

Event-Adjectives
id
event_id
object_id
adjective_id

3. Sample Example