



Ghigliottina

Pierpaolo Basile

Accesso Intelligente all'Informazione
ed Elab. del Linguaggio Naturale



Ghigliottina

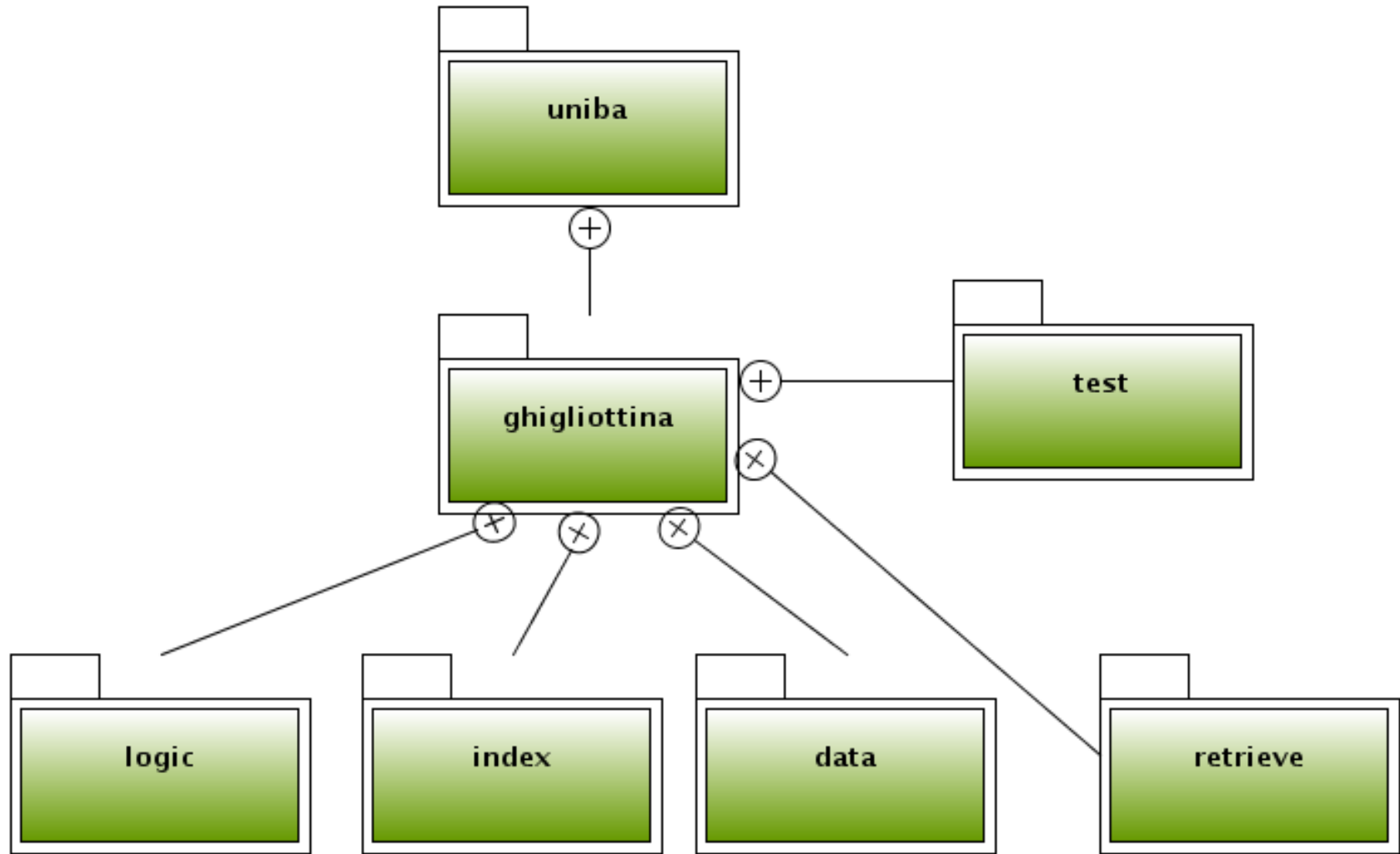
- Valutazione delle nuove sorgenti nel vecchio modello (senza unità cognitive)
- Valutazione delle nuove sorgenti nel nuovo modello
- **NEXT STEP**: modellazione di Wikipedia
 - Valutazione nel nuovo e nel vecchio modello



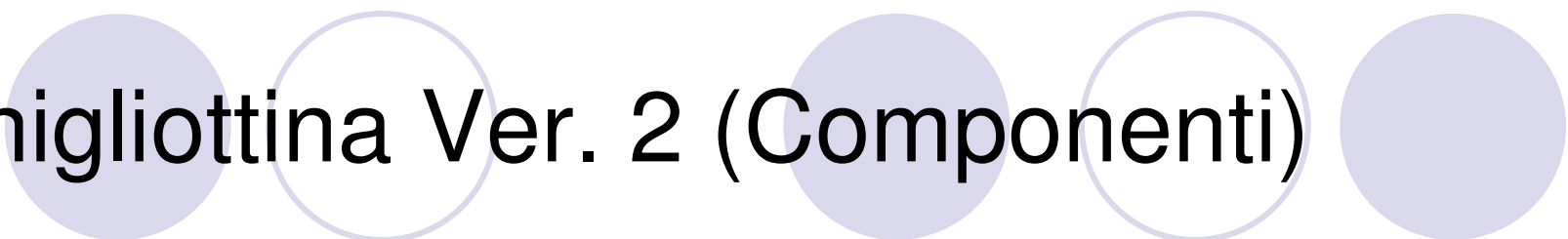
Modellazione di WIKIPEDIA

- Modellazione nel vecchio modello
 - Matrice termini-termini
 - File indice
- Modellazione nel nuovo modello
 - Identificare l'unità cognitiva (UC)
 - Definire la testa e il corpo dell'UC
 - Definire un misura di correlazione dei termini del corpo
 - necessaria se non si intende utilizzare le metriche di pesatura implementate in Lucene

Ghigliottina Ver. 2



Ghigliottina Ver. 2 (Componenti)



- Logic

- implementa i meccanismi di business logic

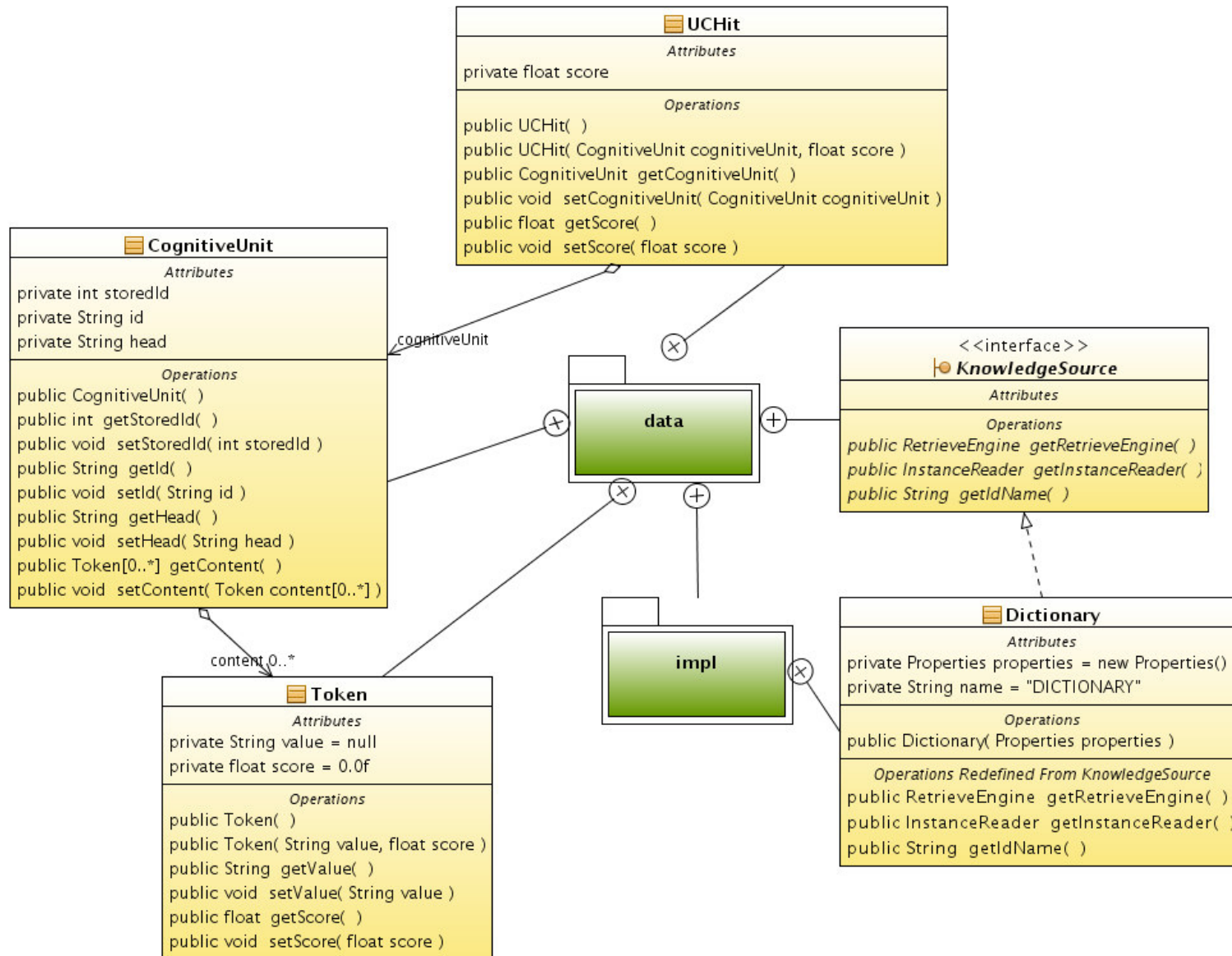
- Index & Retrieve

- implementa i meccanismi di indicizzazione e recupero delle unità cognitive

- Data

- implementa l'astrazione dei dati

Data

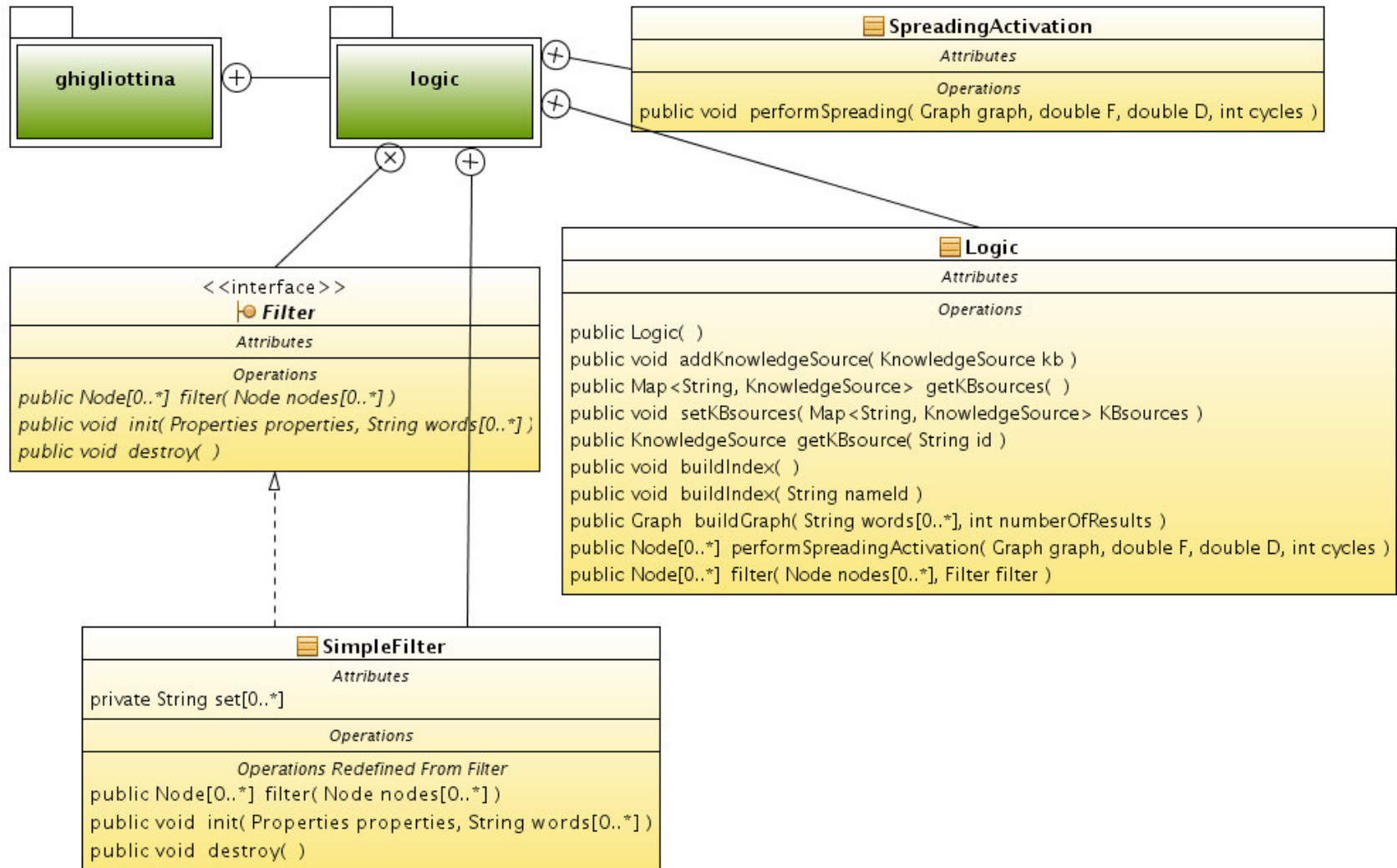


Data

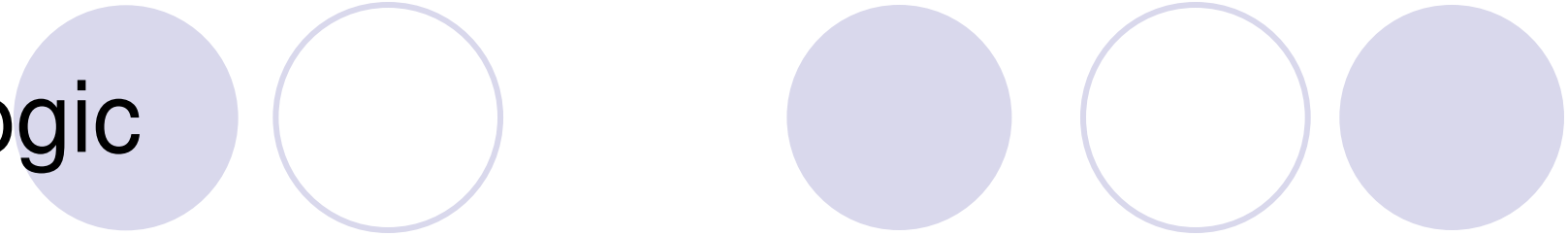


- KnowledgeSource (interfaccia)
 - possiede un id, un motore di retrieval e un motore delle istanze
 - Dictionary: implementa l'interfaccia KnowledgeSource e rappresenta la sorgente dizionario
- CognitiveUnit
 - id, storedId, head, content (una lista di token)
- Token
 - <value,score>
- CUhit: cognitive unit HIT restituito dal motore di retrieval in fase di ricerca

Logic

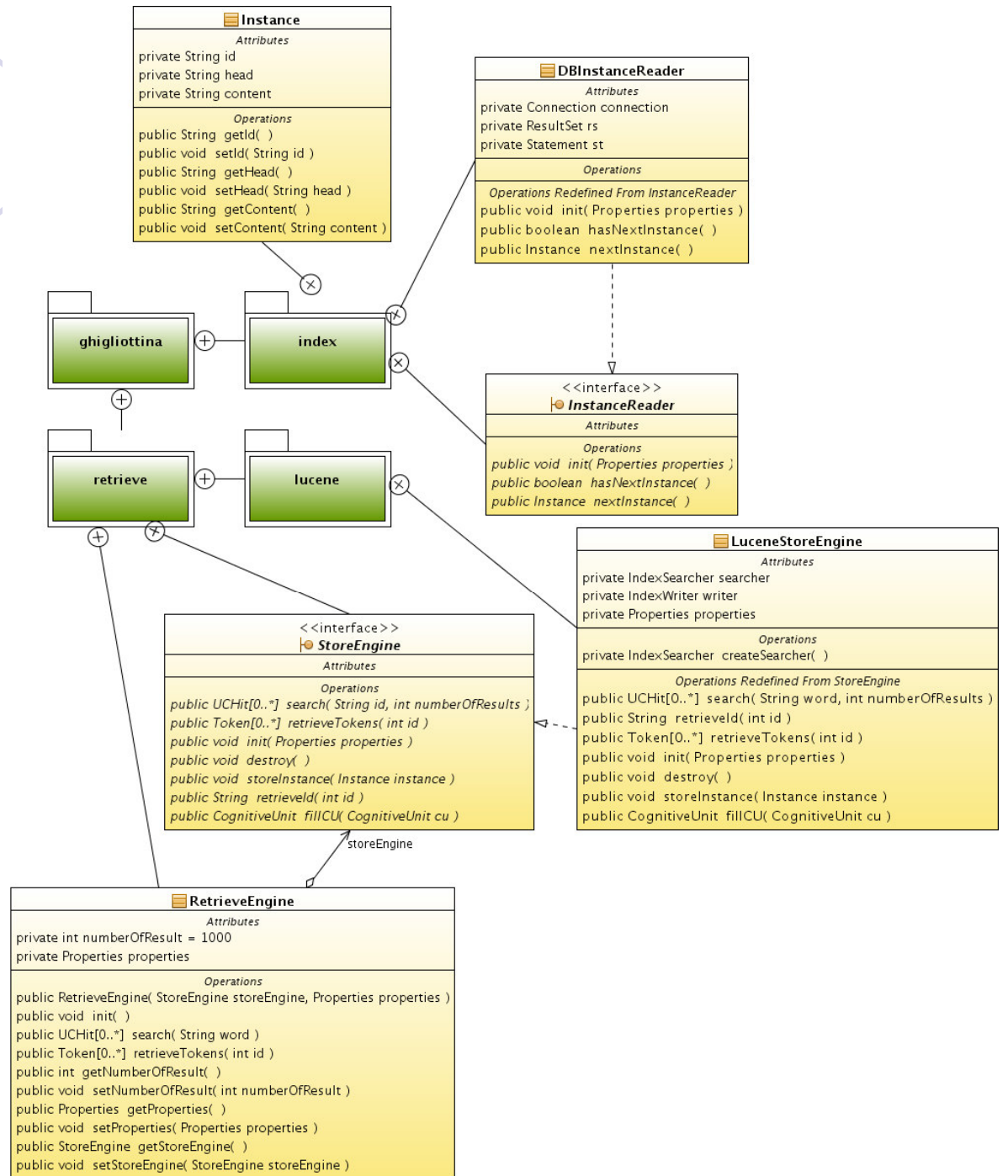


Logic

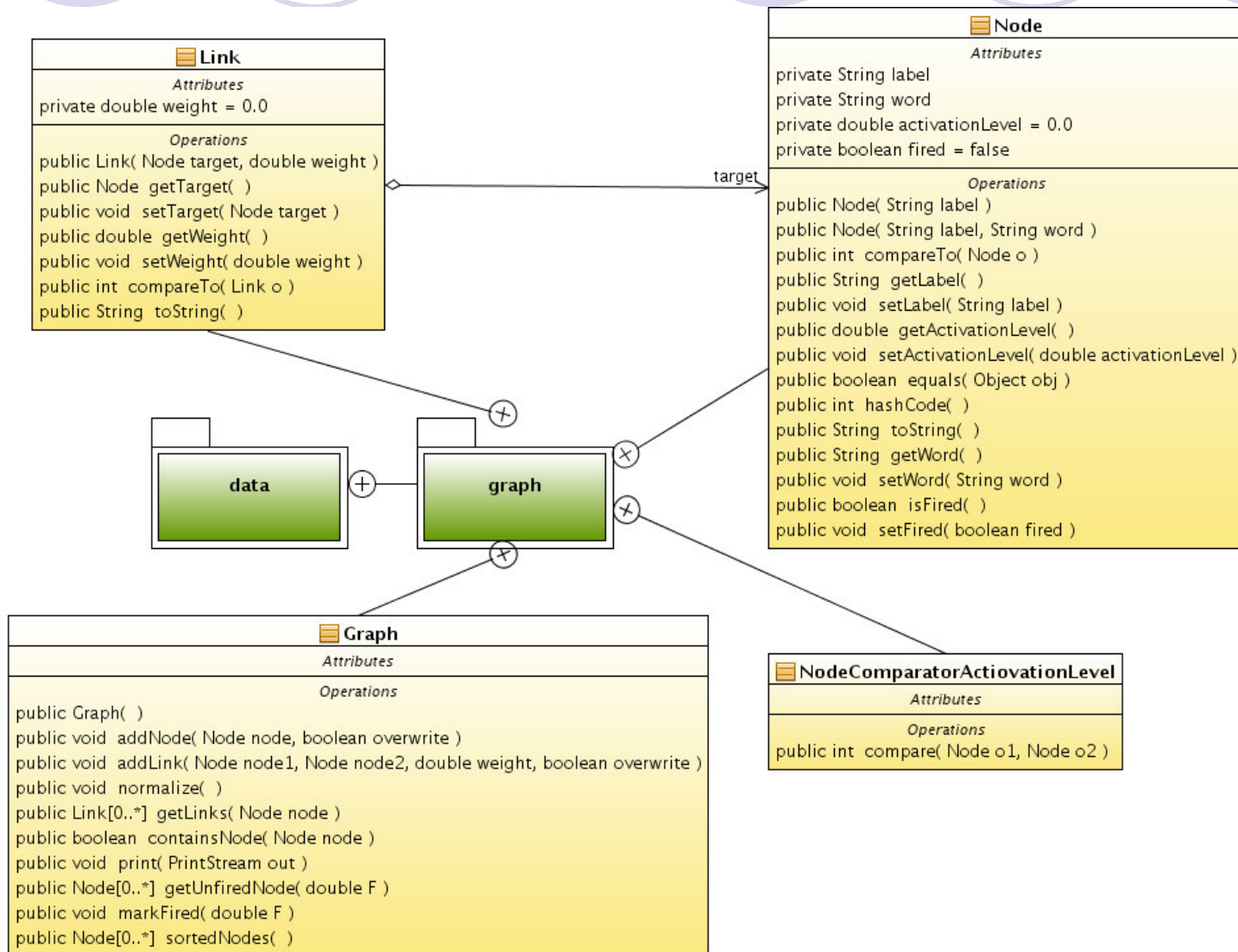


- Filter: rappresenta l'interfaccia per la realizzazione dei filtri
 - Un filtro serve ad eliminare dal result set iniziale delle soluzioni non valide
 - SimpleFilter
 - Implementa l'interfaccia Filter
 - Elimina le soluzioni che contengono come parola gli indizi
- Implementa lo spreading activation

Index and Retrieve



Graph



A decorative header consisting of a row of six circles. The first circle is solid light purple and contains the text 'TO DO'. The second circle is an outline. The third circle is solid light purple. The fourth circle is an outline. The fifth circle is solid light purple. The sixth circle is an outline.

TO DO

- **STEP 1**

- Definire la Cognitive Unit (head, content)
- Definire lo score dei termini estratti nel content
- Processare e salvare

- **STEP 2**

- Realizzare un InstanceReader per leggere le istanze
- Realizzare un motore di storage & retrieval

Agenda



- 22/01/2009

- Consegna delle singole risorse

- Wikipedia TASK

- 27/01/2009: ogni gruppo presenta le sue proposte sulla definizione di UC
- 29/01/2009: ogni gruppo presenta le sue proposte per la pesatura dei termini
- 10/02/2009: consegna del materiale riguardante STEP 1
- 17/02/2009: consegna del materiale riguardante STEP 2

Note

- **Materiale STEP 1**

- Relazione che spieghi le scelte fatte

- Risorsa processata e salvata

- Oltre al formato scelto bisogna consegnare tutti i dati anche nel formato necessario alla prima versione della ghigliottina

- **Materiale STEP 2**

- Relazione che documenti l'instance reader e i motori di retrieval e storage

- Sviluppo delle componenti

GOOGLE group

- <http://groups.google.com/group/aiinlp2008>