ROMA TRE
UNIVERSITÀ DEGLI STUDI

9th International Workshop on Web Information Systems Modeling (WISM 2012)

# Web Navigation via Semantic Annotations

Roberto De Virgilio and Lorenzo Dolfi

# Scenario


Wikipedia results articles that have scores not greater than 8.3%
Query: Semantic Web Conference

# Syntactic Search

## Search results

From Wikipedia, the free encyclopedia

kenneth iverson apl [Search]

Content pages | Multimedia | Help and Project pages | Everything | Advanced

Search for files

Ken Iverson (redirect from Kenneth Iverson)

Ken **Iverson** can refer to: 2004), developer of the **APL** programming language ... F. **Kenneth Iverson** (1925–2002), former CEO of the Nucor Corporation ...

237 B (29 words) - 16:50, 1 December 2009

WIKIPEDIA
*The Free Encyclopedia*

## Ken Iverson

From Wikipedia, the free encyclopedia

(Redirected from Kenneth Iverson)

**Ken Iverson** can refer to:

- Kenneth E. Iverson (1920–2004), developer of the APL programming language
- F. Kenneth Iverson (1925–2002), former CEO of the Nucor Corporation

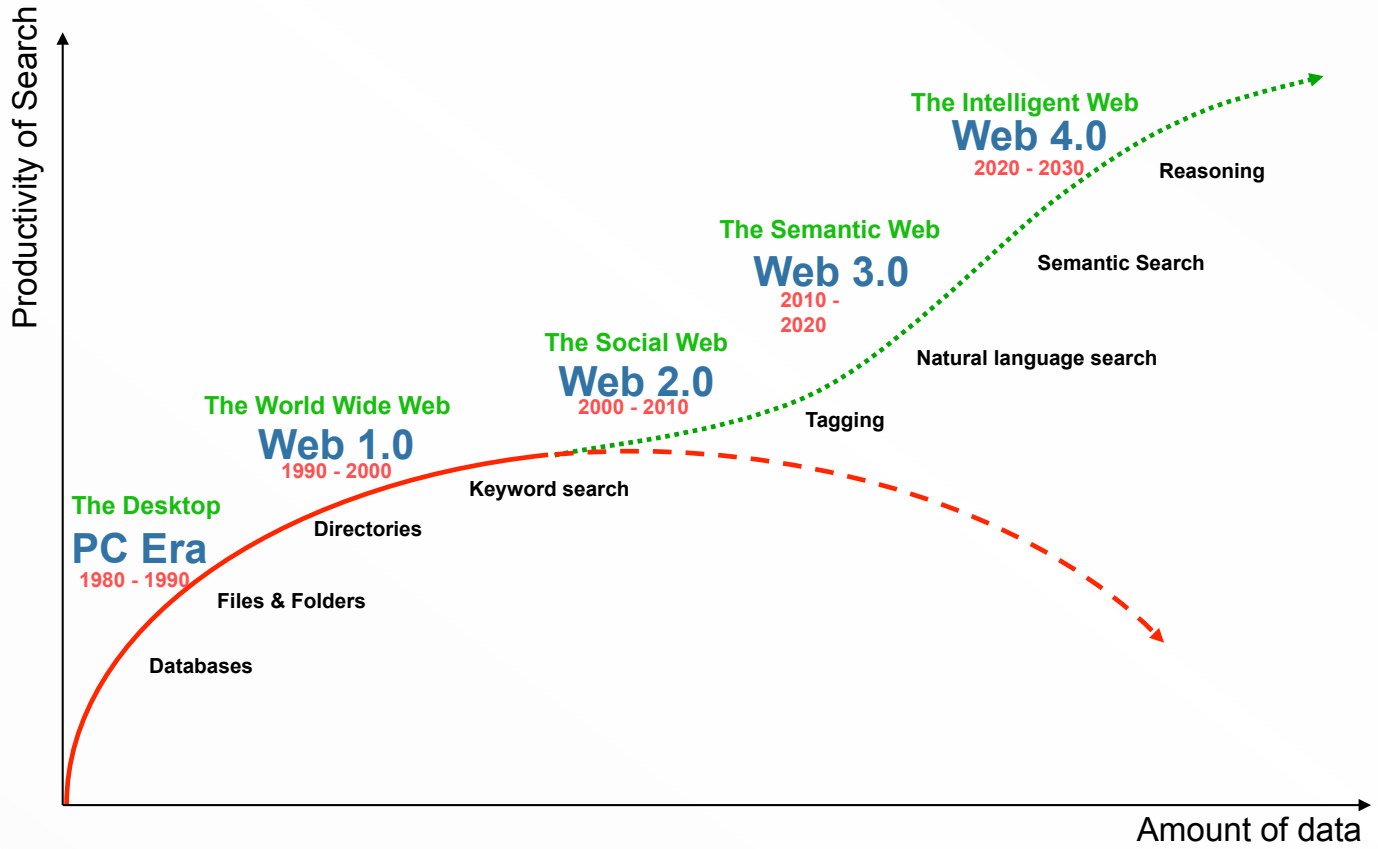
## Kenneth E. Iverson

From Wikipedia, the free encyclopedia

This article includes a list of references, related reading or external links, but **its sources remain unclear** citations. Please improve this article by introducing more precise citations where appropriate. *(May 2009)*

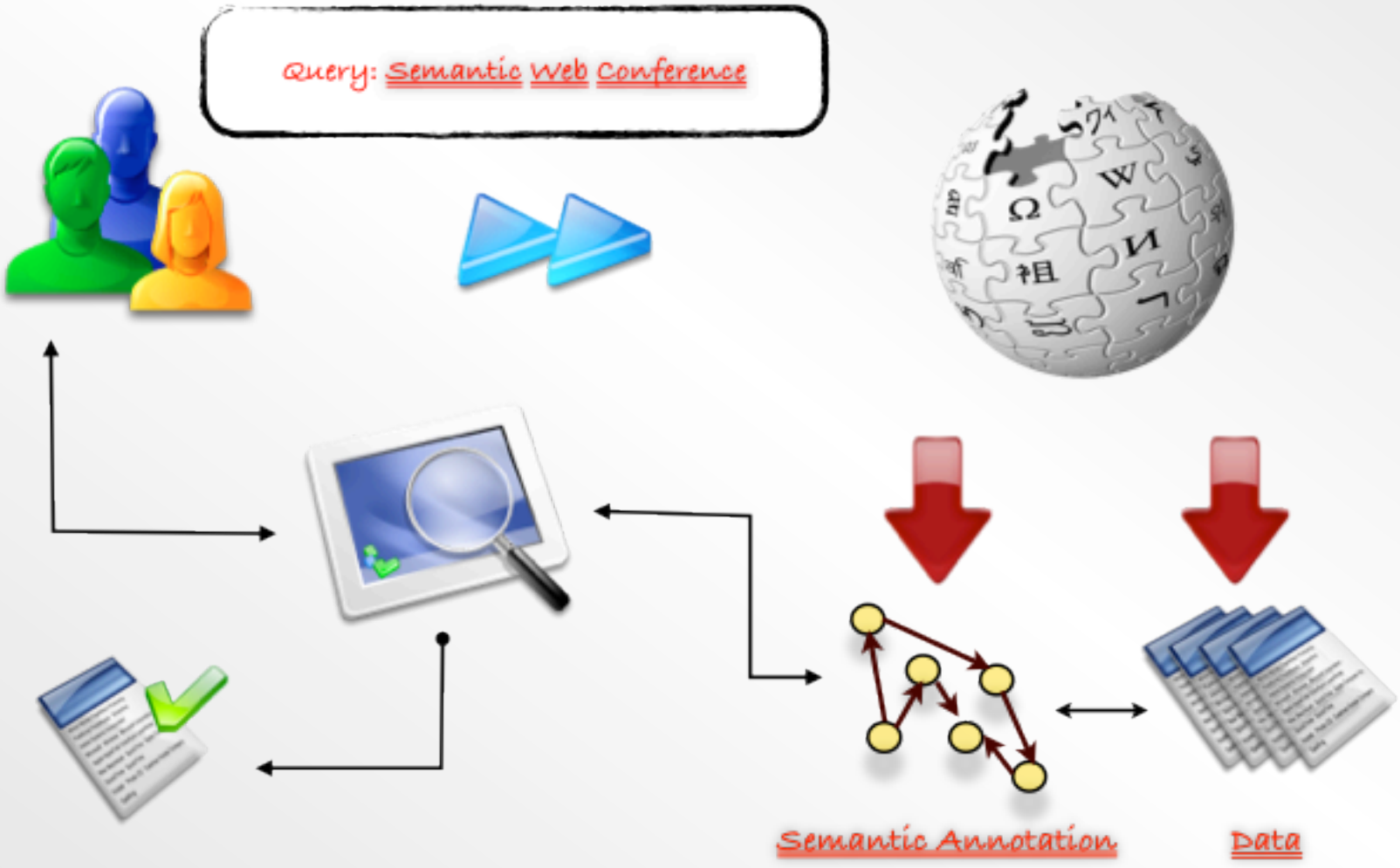
**Kenneth Eugene Iverson** (17 December 1920 - 19 October 2004) was a Canadian computer scientist noted for the development of the APL programming language in 1962. He was honored with the Turing Award in 1979 for his contributions to mathematical notation and programming language theory. The Iverson Award for contributions to APL was named in his honor.

# Beyond the limits of Keyword Search*




*Nova Spivack. Making Sense of the Semantic Web

# Semantic Search



# Semantic Search



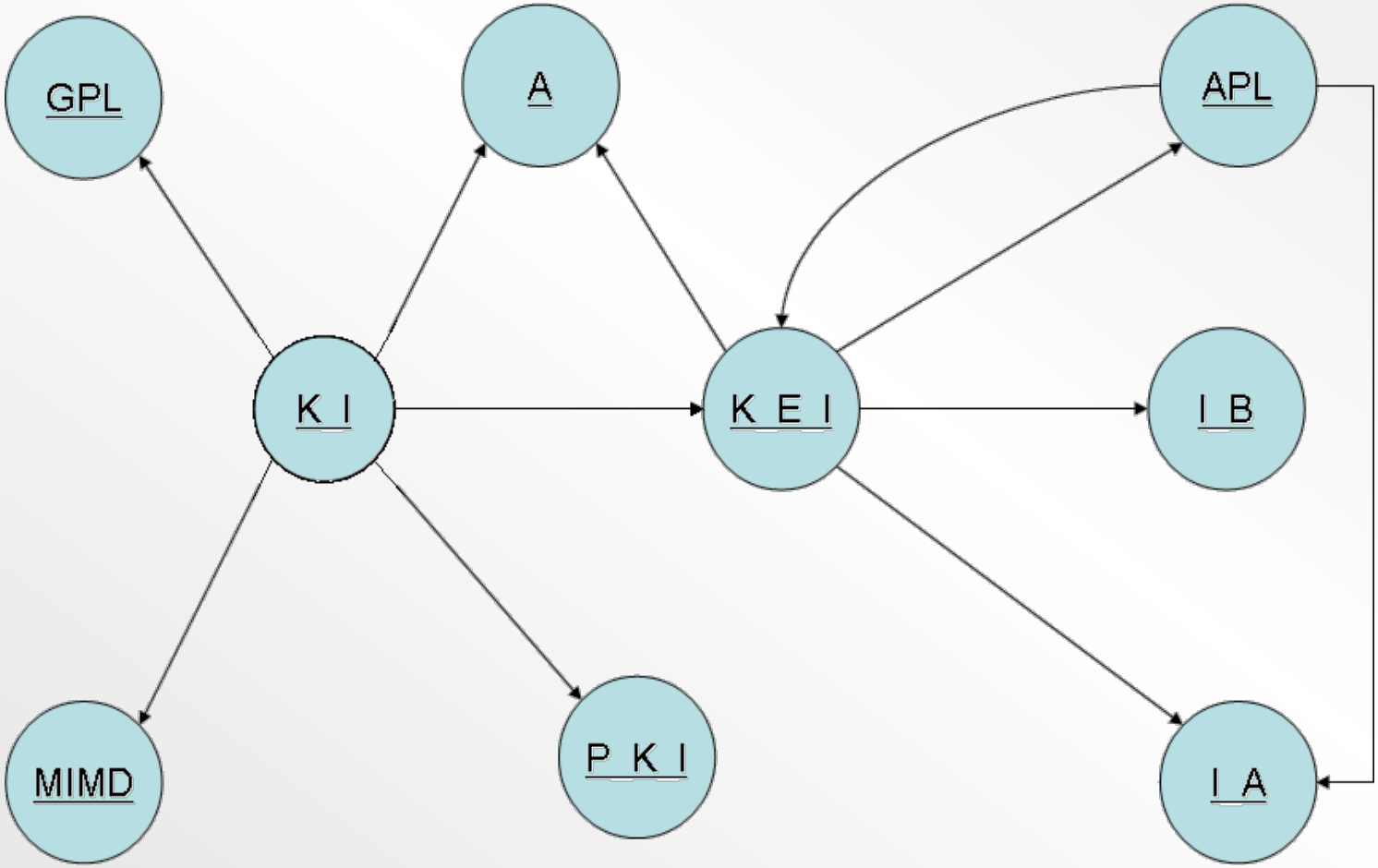




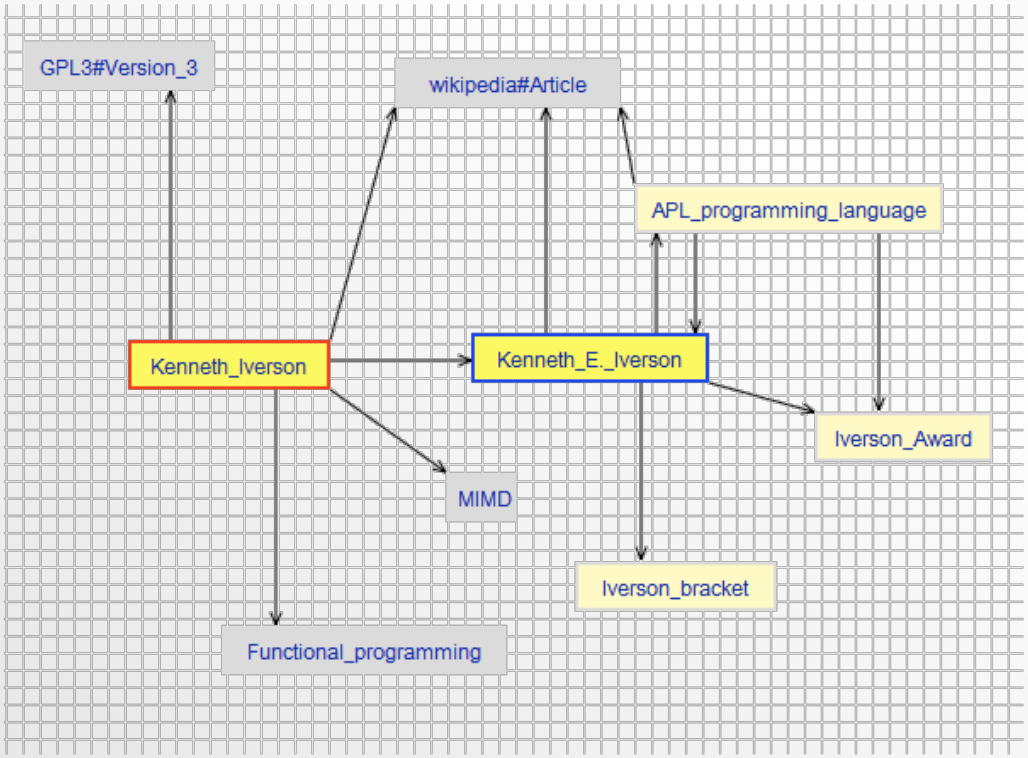
Semantic Search

- *Solution 1*
- *Solution 2*
- .....
- *Solution n*

# Problem



# Proposed Solution

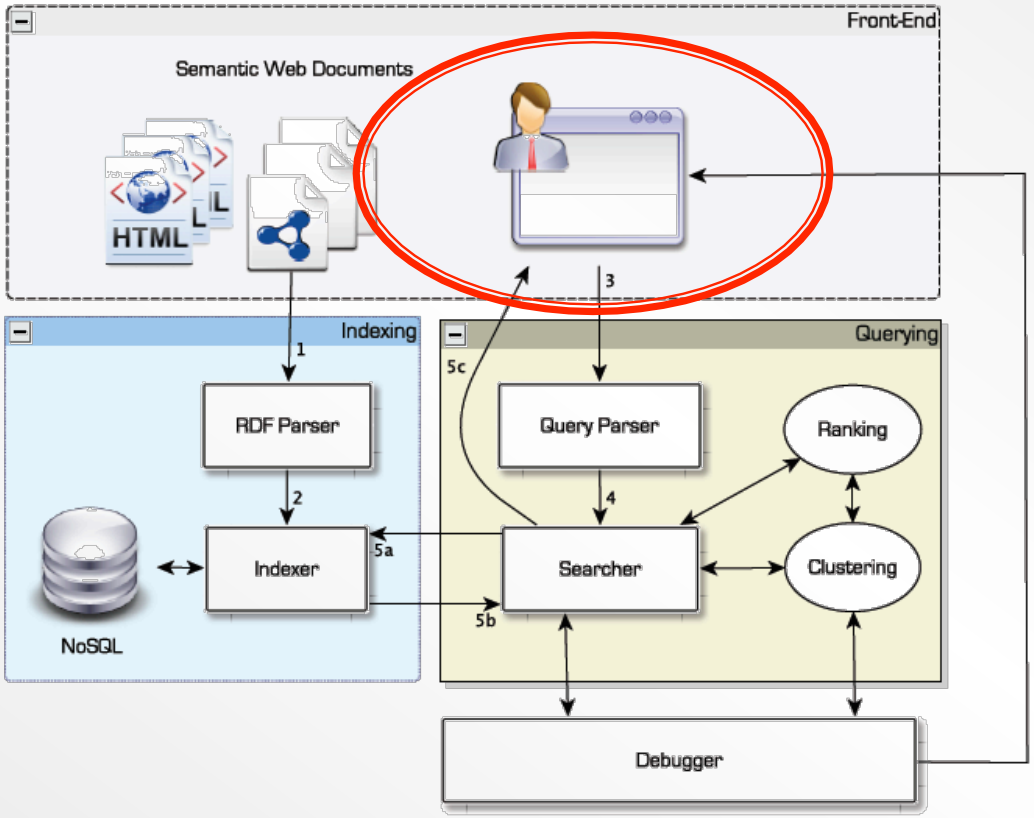



# Goal

***Development of a set of algorithms for an effective navigation of Web pages***

1. Computation of *Semantic Centroids*
2. Computation of *Structural Centroids*
3. Interactive visualization of *Semantic Annotations* vs Web pages

# Yaanii: Keyword Search over RDF



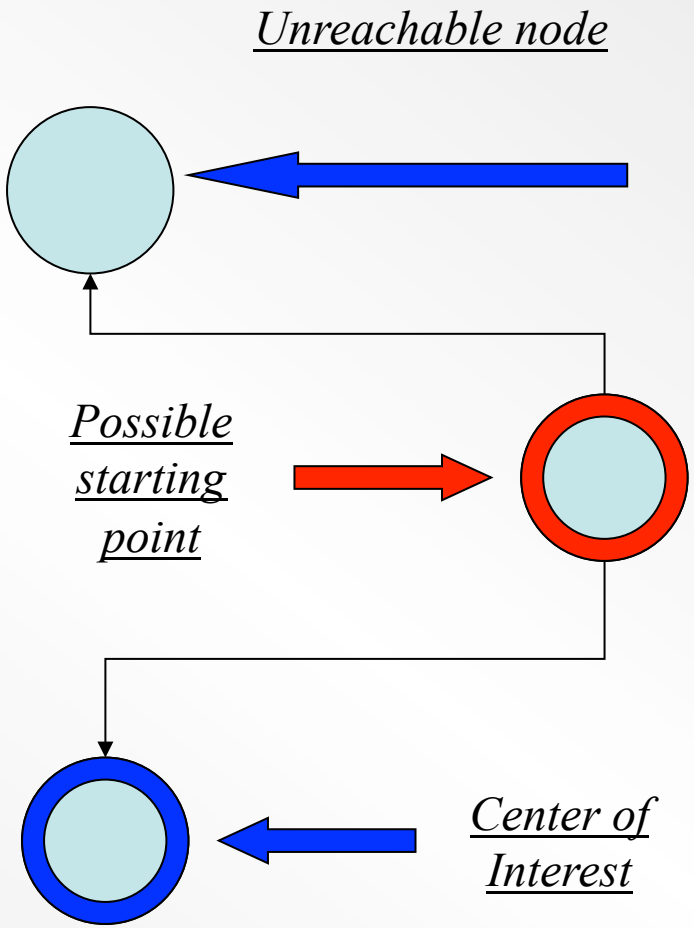

*R. De Virgilio et al. Cluster-based exploration for Effective Keyword Search over Semantic Datasets. ER 2009.

**R. De Virgilio et al. A Path-Oriented RDF Index for Keyword Search Query Processing. DEXA 2011.

# Structural and Semantic Centers

## Two kinds of center

1. It is possible that the center of interest is not well connected with all other nodes; some of them could be unreachable.
2. Finding a starting point could be very tedious
3. The computed center of interest could not be the wanted one

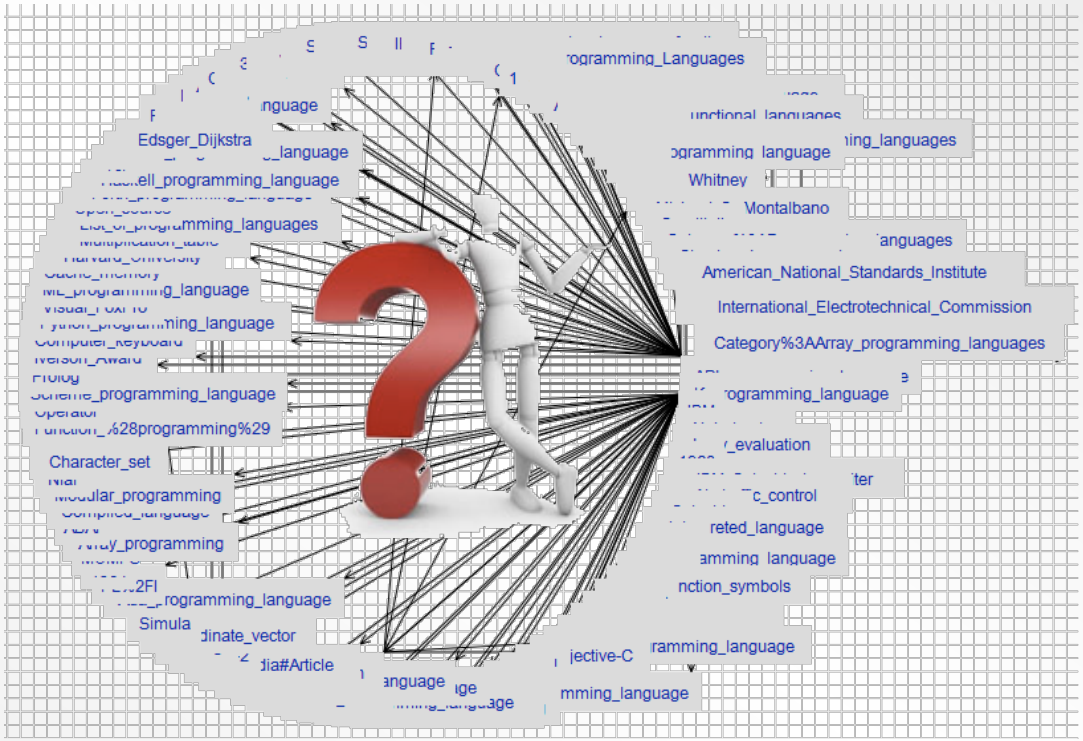


# Structural Center of the Graph

## *Providing a point to start navigation*

Three different algoritms:

- ***Eccentricity***
- ***Closeness***
- ***Centroid***




*Let us guarantee to reach as many nodes as possible*

# Eccentricity - Meaning

*Minimize the maximum distance*

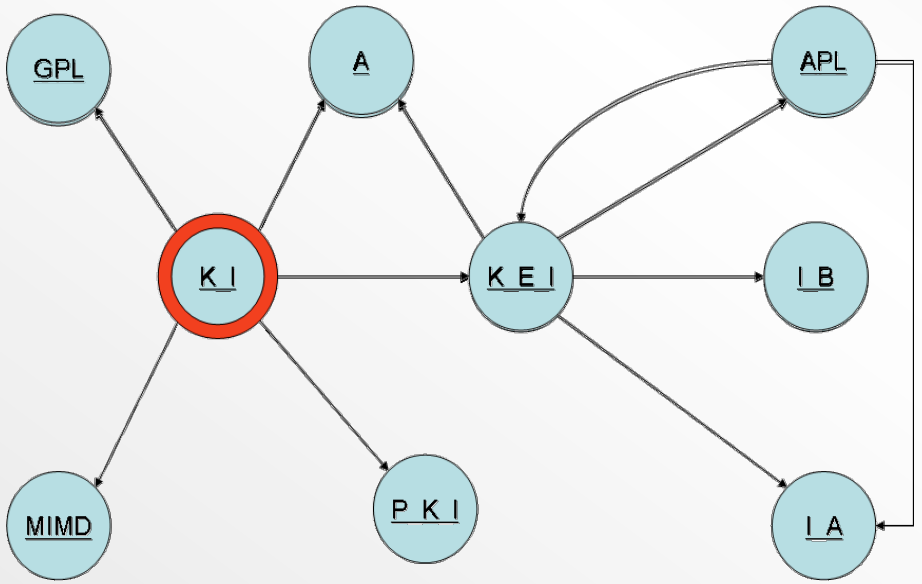
*Finding the node in the central position of the graph*

1. For each node is computed the maximum distance ***d*** from the other nodes
2. The center will be the node with the minimum ***d***

The elected node ensures a max of ***d*** steps to reach any other node.

***d*** = *Radius of the Graph*

# Eccentricity - Execution




$$M = \begin{bmatrix} 0 & \infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty \\ \infty & 0 & \infty & \infty & \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & 0 & \infty & \infty & \infty & \infty & \infty & \infty \\ 1 & 1 & 1 & 0 & 1 & 1 & 2 & 2 & 2 \\ \infty & \infty & \infty & \infty & 0 & \infty & \infty & \infty & \infty \\ \infty & \infty & 1 & \infty & \infty & 0 & 1 & 1 & 1 \\ \infty & \infty & 2 & \infty & \infty & 1 & 0 & 2 & 1 \\ \infty & \infty & \infty & \infty & \infty & \infty & \infty & 0 & \infty \\ \infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty & 0 \end{bmatrix}$$

*idx(GPL) = 1, idx(MIMD) = 2,*
*idx(A) = 3, idx(KI) = 4,*
*idx(PKI) = 5, idx(KEI) = 6,*
*idx(APL) = 7, idx(IB) = 8 idx*
*(IA) = 9*

$$L_E = \begin{bmatrix} \infty & \infty & \infty & 2 & \infty & \infty & \infty & \infty & \infty \end{bmatrix}^t$$

# Closeness - Meaning

*Minimize the sum of distances*

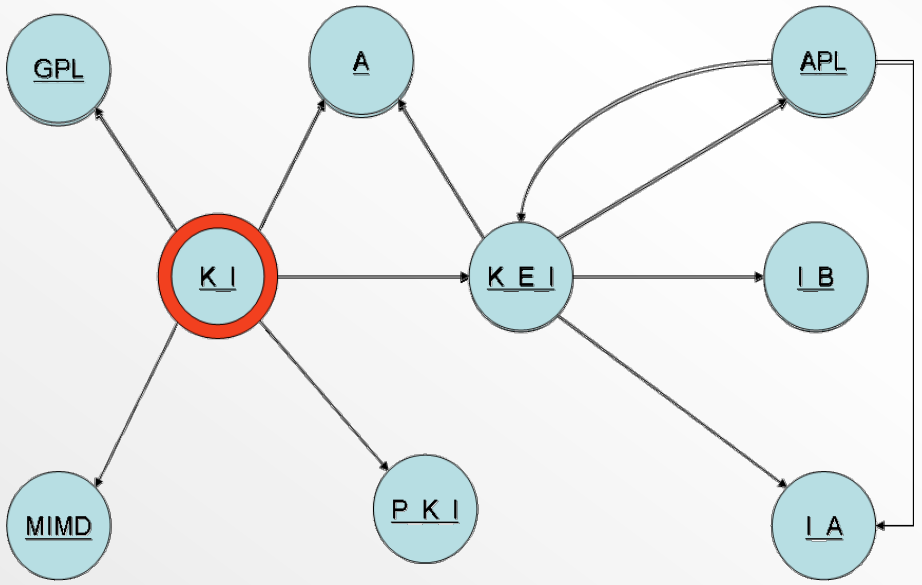
*Finding the best integrated node in the graph*

1. For each node we compute the sum ***s*** of the distances from all the other nodes

2. The node with the minimum ***s*** is elected as structural center

A well integrated node in a graph has many links with other nodes.

# Closeness - Execution




$$M = \begin{bmatrix} 0 & \infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty \\ \infty & 0 & \infty & \infty & \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & 0 & \infty & \infty & \infty & \infty & \infty & \infty \\ 1 & 1 & 1 & 0 & 1 & 1 & 2 & 2 & 2 \\ \infty & \infty & \infty & \infty & 0 & \infty & \infty & \infty & \infty \\ \infty & \infty & 1 & \infty & \infty & 0 & 1 & 1 & 1 \\ \infty & \infty & 2 & \infty & \infty & 1 & 0 & 2 & 1 \\ \infty & \infty & \infty & \infty & \infty & \infty & \infty & 0 & \infty \\ \infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty & 0 \end{bmatrix}$$

*idx(GPL) = 1, idx(MIMD) = 2,*
*idx(A) = 3, idx(KI) = 4,*
*idx(PKI) = 5, idx(KEI) = 6,*
*idx(APL) = 7, idx(IB) = 8 idx*
*(IA) = 9*

$$L_C = \begin{bmatrix} \infty & \infty & \infty & 11 & \infty & \infty & \infty & \infty & \infty \end{bmatrix}^t$$

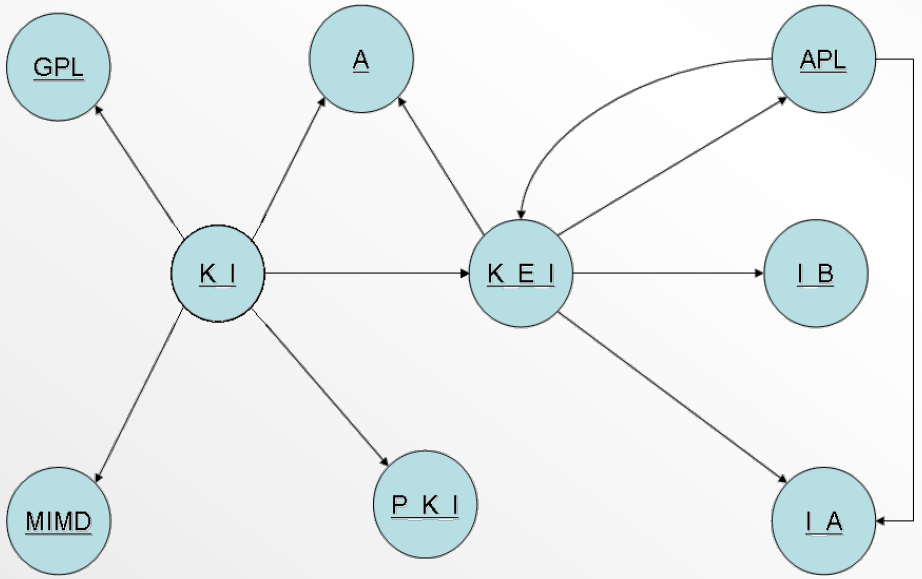
# Centroid - Meaning

*Maximize the "convenience"*

*Finding the node linked to the majority of nodes as possible*

Maximizes the probability that the structural center is as close as possible to the semantic one, however, without limiting freedom of navigation

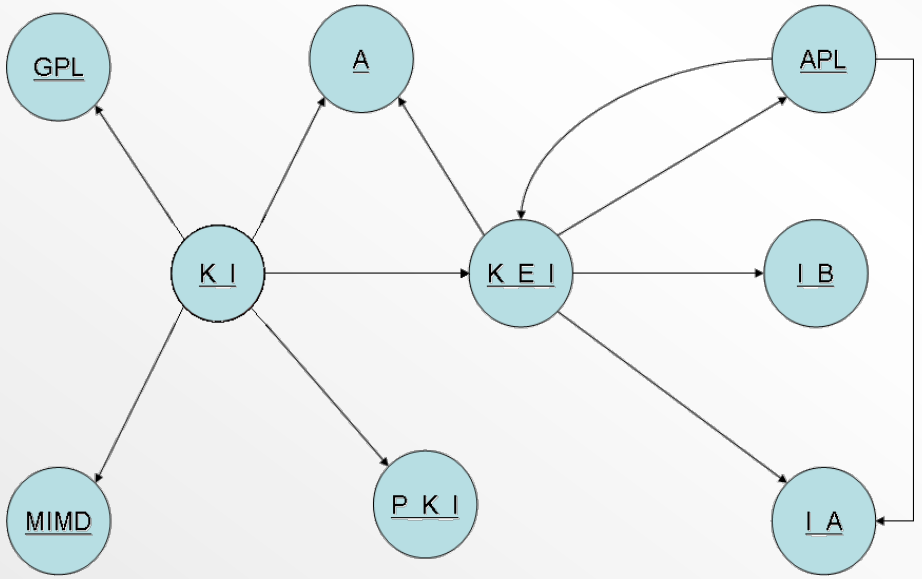
# Centroid - Execution




$$M = \begin{bmatrix} 0 & \infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty \\ \infty & 0 & \infty & \infty & \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & 0 & \infty & \infty & \infty & \infty & \infty & \infty \\ 1 & 1 & 1 & 0 & 1 & 1 & 2 & 2 & 2 \\ \infty & \infty & \infty & \infty & 0 & \infty & \infty & \infty & \infty \\ \infty & \infty & 1 & \infty & \infty & 0 & 1 & 1 & 1 \\ \infty & \infty & 2 & \infty & \infty & 1 & 0 & 2 & 1 \\ \infty & \infty & \infty & \infty & \infty & \infty & \infty & 0 & \infty \\ \infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty & 0 \end{bmatrix}$$

*idx(GPL) = 1, idx(MIMD) = 2, idx(A) = 3, idx(KI) = 4, idx(PKI) = 5,*
*idx(KEI) = 6, idx(APL) = 7, idx(IB) = 8 idx(IA) = 9*

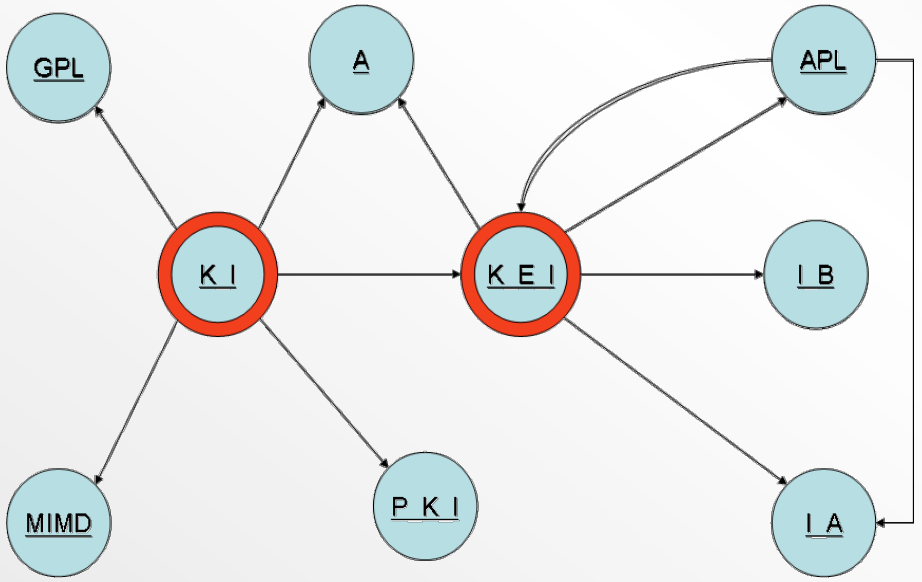
# Centroid - Execution




$$C = \begin{bmatrix} \infty & & & & & & & & \\ & \infty & & & & & & & \\ & & \infty & & & & & & \\ & & & \infty & & & & & \\ & & & & \infty & & & & \\ & & & & & \infty & & & \\ & & & & & & \infty & & \\ & & & & & & & \infty & \\ & & & & & & & & \infty \end{bmatrix}$$

*idx(GPL) = 1, idx(MIMD) = 2, idx(A) = 3, idx(KI) = 4, idx(PKI) = 5,*
*idx(KEI) = 6, idx(APL) = 7, idx(IB) = 8 idx(IA) = 9*

# Centroid - Execution




$$C = \begin{bmatrix} \infty & 0 & 0 & -7 & 0 & -4 & -4 & 0 & 0 \\ 0 & \infty & 0 & -7 & 0 & -4 & -4 & 0 & 0 \\ 0 & 0 & \infty & -7 & 0 & -4 & -4 & 0 & 0 \\ 7 & 7 & 7 & \infty & 7 & 0 & 3 & 7 & 7 \\ 0 & 0 & 0 & -7 & \infty & -4 & -4 & 0 & 0 \\ 4 & 4 & 4 & 0 & 4 & \infty & 2 & 4 & 4 \\ 4 & 4 & 4 & -3 & 4 & -2 & \infty & 4 & 4 \\ 0 & 0 & 0 & -7 & 0 & -4 & -4 & \infty & 0 \\ 0 & 0 & 0 & -7 & 0 & -4 & -4 & 0 & \infty \end{bmatrix}$$

*idx(GPL) = 1, idx(MIMD) = 2,*
*idx(A) = 3, idx(KI) = 4,*
*idx(PKI) = 5, idx(KEI) = 6,*
*idx(APL) = 7, idx(IB) = 8 idx*
*(IA) = 9*

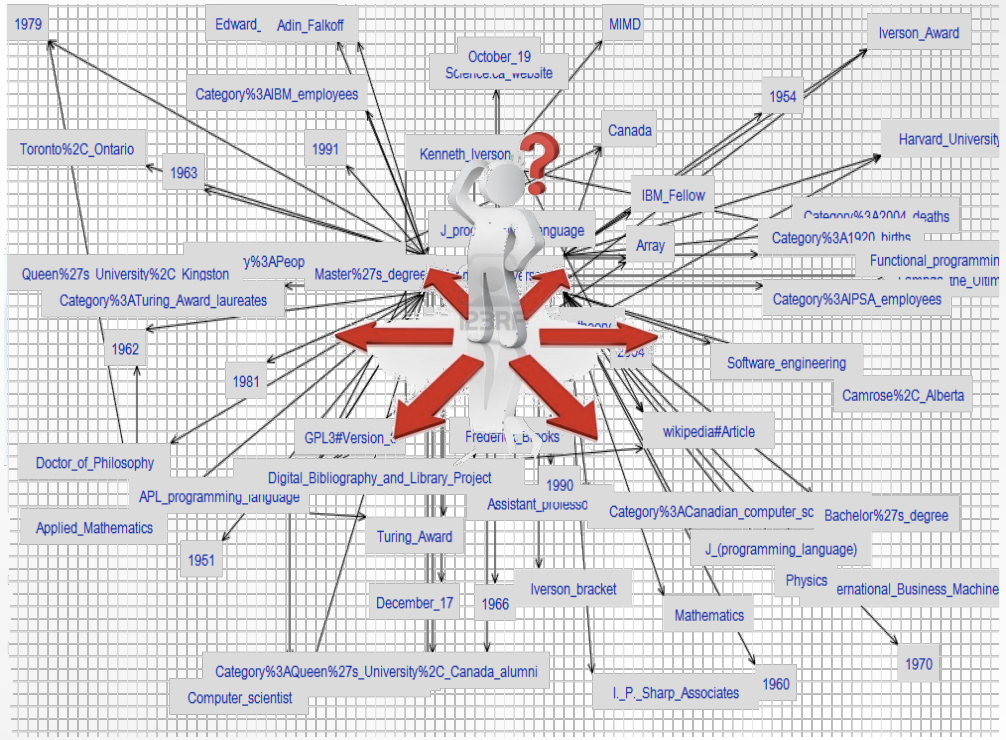
$$min = \begin{bmatrix} -7 & -7 & -7 & 0 & -7 & 0 & -3 & -7 & -7 \end{bmatrix}^t$$

ROMA TRE
UNIVERSITÀ DEGLI STUDI

# Semantic Center of the Graph

## *Providing the center of interest of the search*

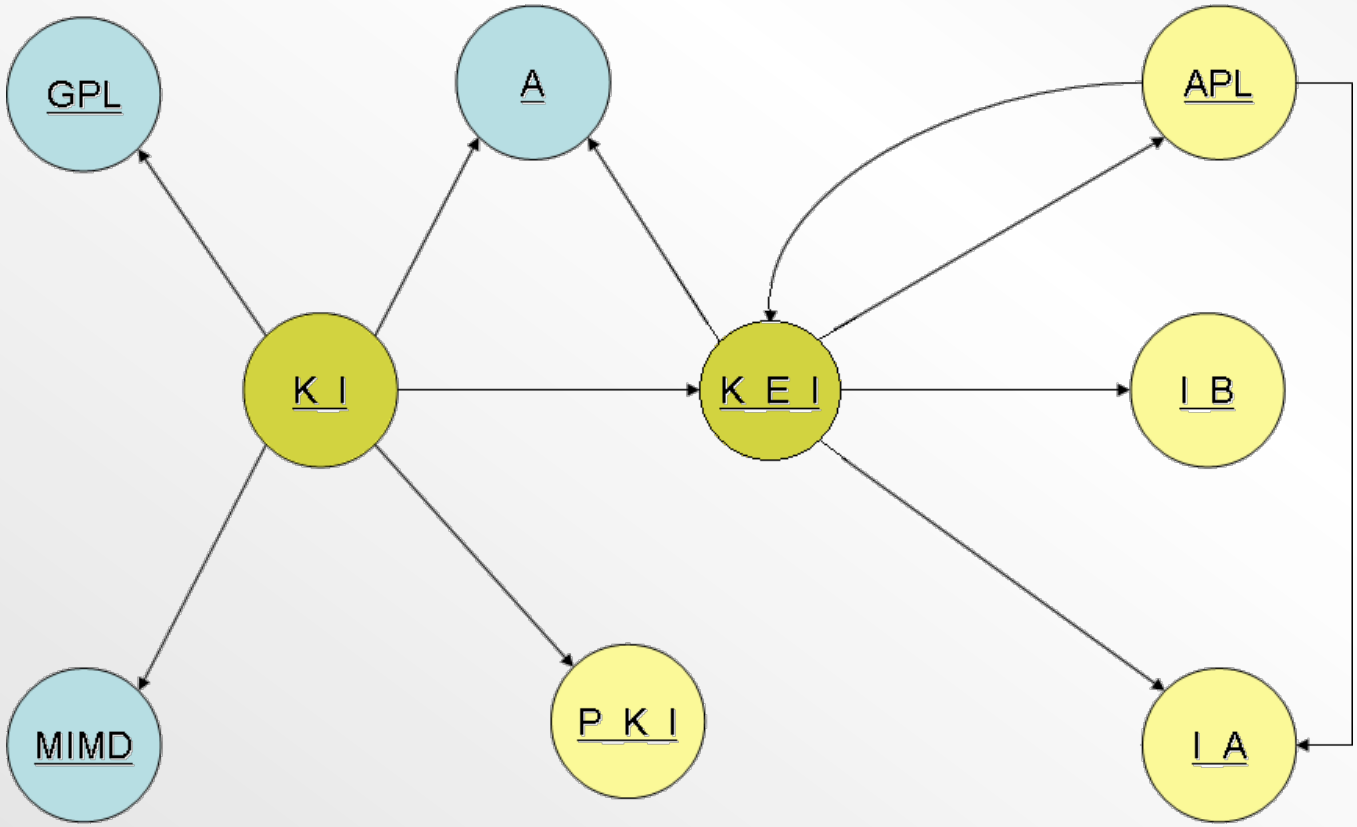
- Provides a direction to follow
- Represents the *heart of the research*
- Computation is based on the *Keywords*




*Let us guarantee both accuracy and freedom of navigation*

# Reference Graph

Query: "***KENNETH IVERSON APL***"




**MATCHES**

**A = 0**
**GPL = 0**
**MIMD = 0**
**PKI = 1**
**IA = 1**
**IB = 1**
**APL = 1**
**KI = 2**
**KEI = 2**

# Barycenter - Meaning

*Minimize the sum of distances*

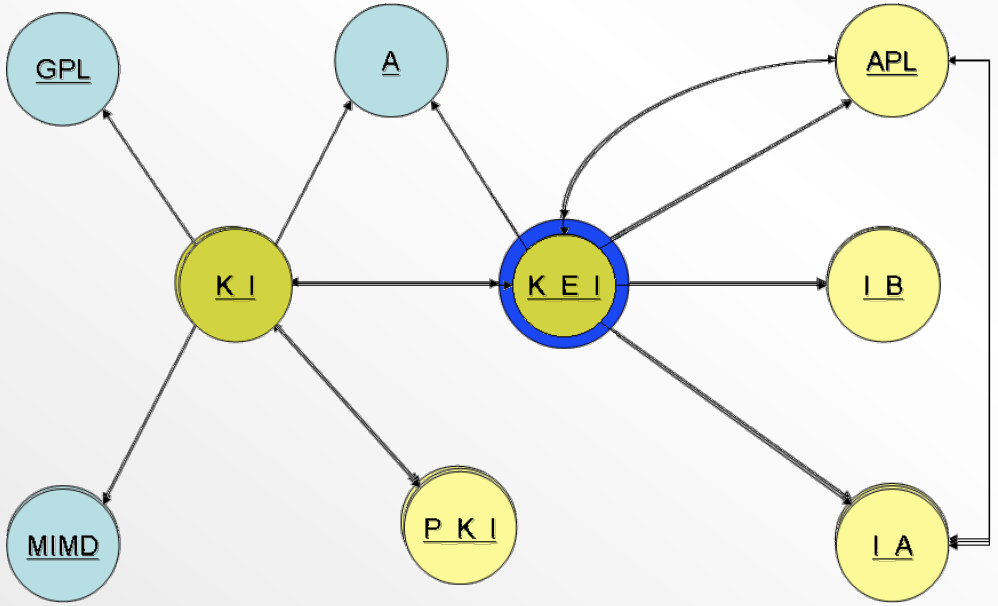
*Finding the best integrated node in a **semantic** graph*

1. Only nodes with *rank* > 0 are considered
2. Arcs are considered as non-oriented
3. Closeness is applied and then the result is weighted by the *rank* of the nodes.

Computes the best integrated node (*Closeness*) in a graph, whose arcs are conceptual links (*semantic*) between the resources.

Physical structure is not considered

# Barycenter - Execution

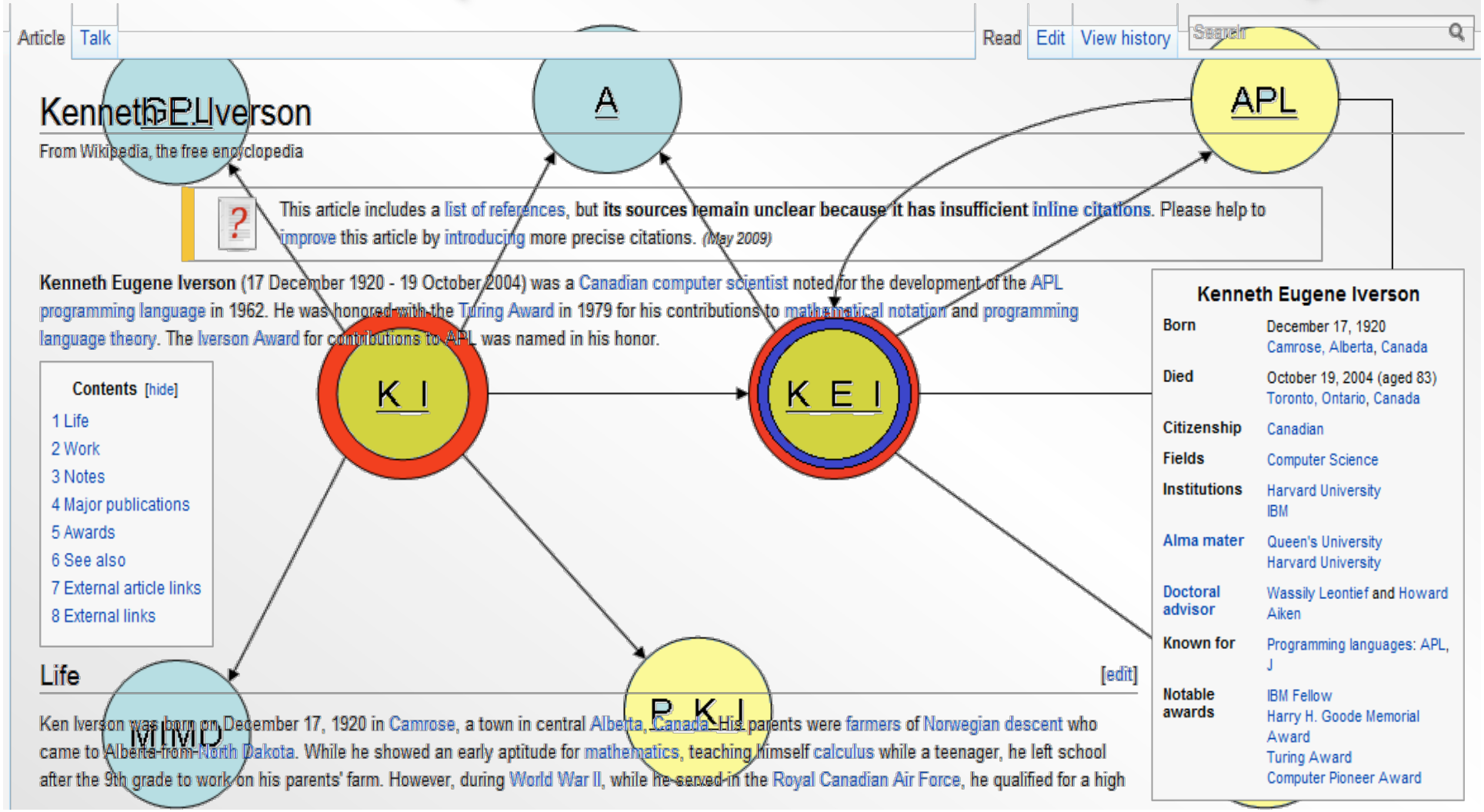



$$\begin{bmatrix} 0 & 1 & 1 & 2 & 2 & 2 \\ 1 & 0 & 2 & 3 & 3 & 3 \\ 1 & 2 & 0 & 1 & 1 & 1 \\ 2 & 3 & 1 & 0 & 2 & 1 \\ 2 & 3 & 1 & 2 & 0 & 2 \\ 2 & 3 & 1 & 1 & 2 & 0 \end{bmatrix}$$

*idx(GPL) = 1, idx(MIMD) = 2, idx(A) = 3, idx(KI) = 4, idx(PKI) = 5, idx(KEI) = 6, idx(APL) = 7, idx(IB) = 8 idx (IA) = 9*

$$D = \left[\infty \; \infty \; \infty \; \frac{8}{2} \; \frac{12}{1} \; \frac{6}{2} \; \frac{9}{1} \; \frac{10}{1} \; \frac{9}{1}\right]^t$$

# centers (structural and semantic)



# Implementation

- Interaction with *semantic research algorithms*

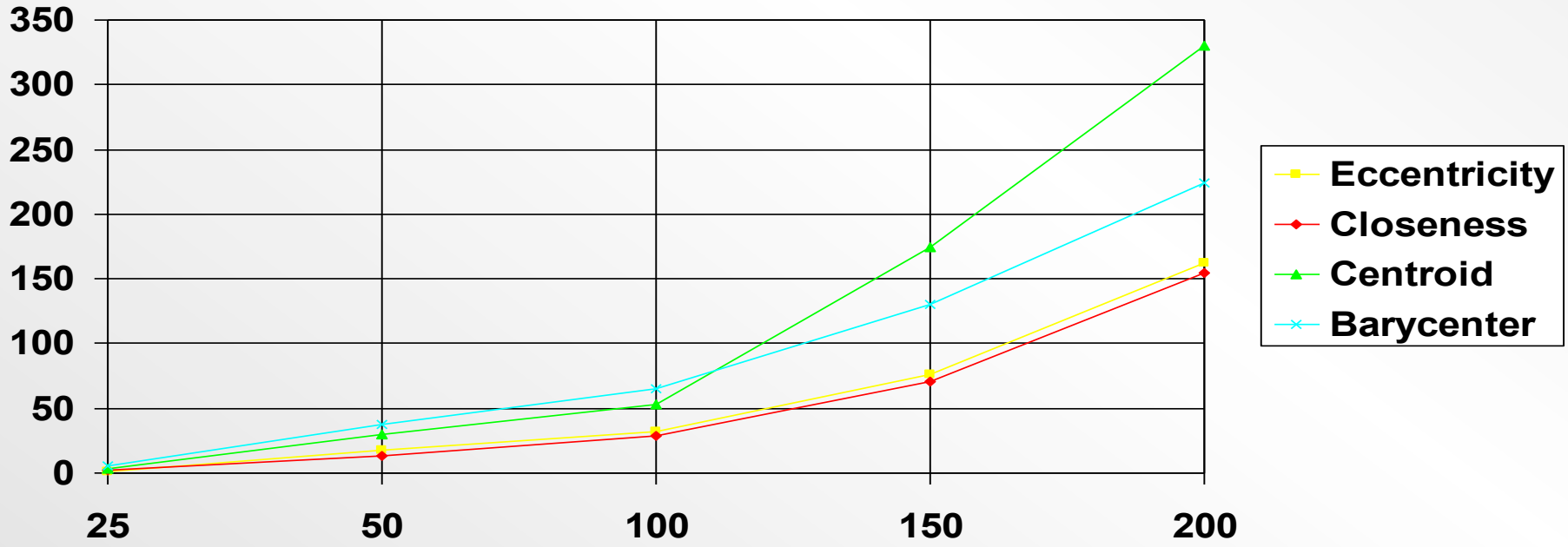
Yaanii
Sama

## *Google Web Toolkit*

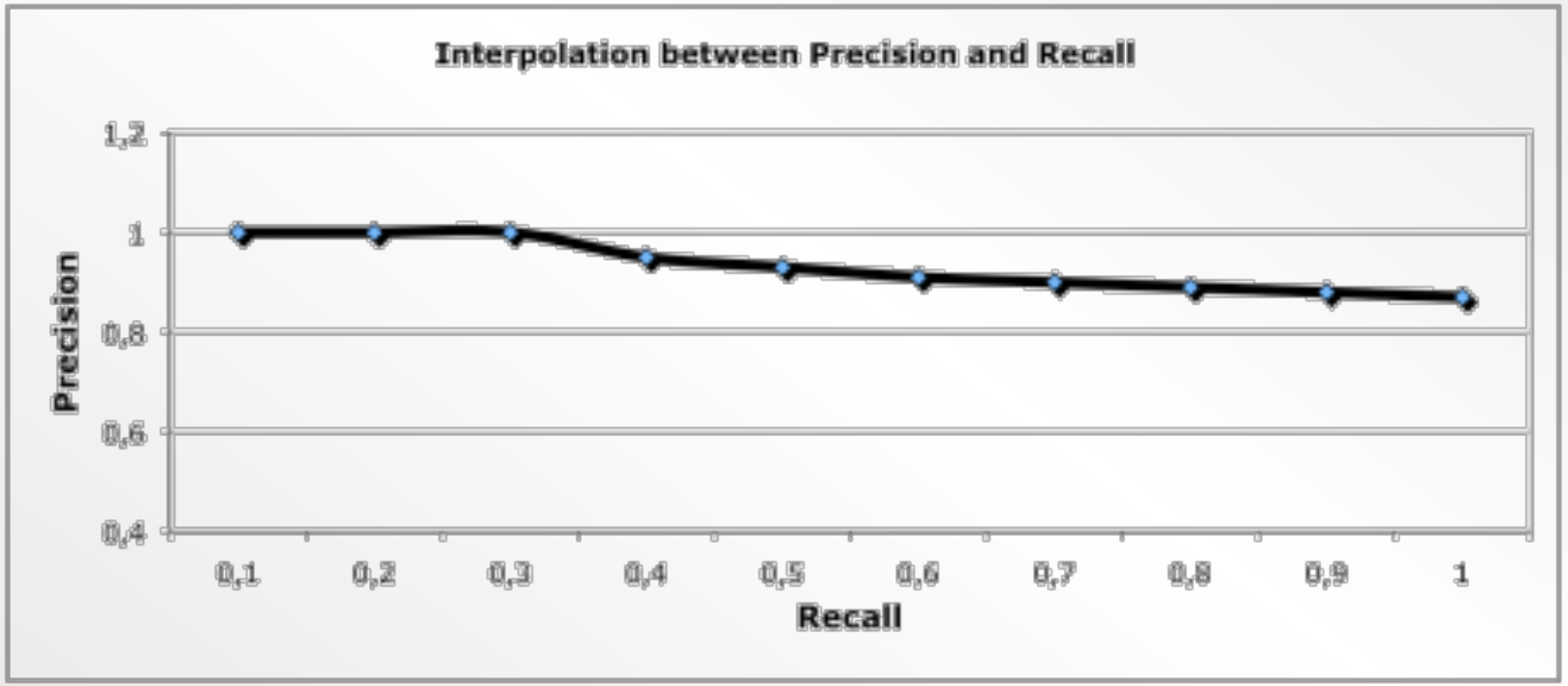
- Code translation from *Java* to *JavaScript*
- Asynchronous calls according to the *Ajax* design style

# Tests

| | 25 | 50 | 100 | 150 | 200 |
|---|---|---|---|---|---|
| Eccentricity | 1,5874 | 17,213 | 31,9694 | 75,9605 | 162,8158 |
| Closeness | 1,6818 | 13,6584 | 28,954 | 70,121 | 154,2781 |
| Centroid | 3,6147 | 29,5762 | 53,1725 | 174,4683 | 330,1163 |
| Barycenter | 5,1862 | 37,0458 | 65,1507 | 129,9 | 224,4051 |



# Tests


Interpolation between Precision and Recall
Precision
1,2
1
0,8
0,6
0,4
0,1
0,2
0,3
0,4
0,5
0,6
0,7
0,8
0,9
1
Recall

# Conclusion and Future Work

*Interactive* visualization of semantic annotations for an **effective** navigation of Web pages


- *Optimization* of the proposed algorithms
- Collect **statistics** to score the nodes of the RDF graph
- Implementation trought *Web Services* and *Silverlight*.

# Thanks for the attention