

Looking for Experts? What can Linked Data do for you?

Milan Stankovic ♦ Claudia Wagner ♦ Jelena Jovanovic ♦ Philippe Laublet



What's the problem of traditional expert search systems?

structured data from
closed systems



Closed-world view



unstructured data
from open systems



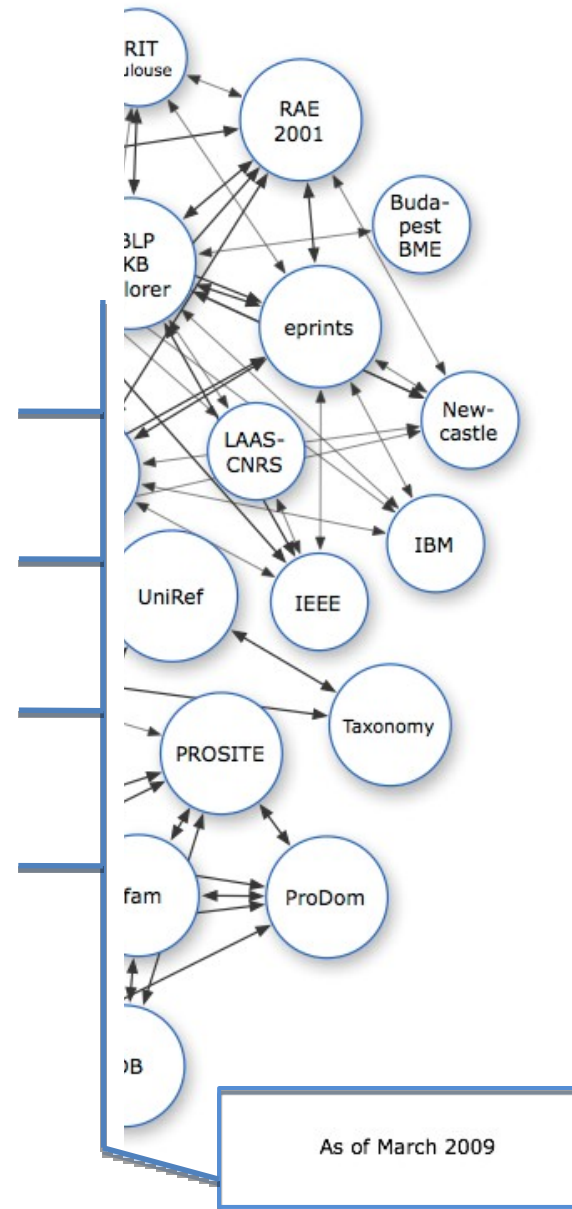
Lack of semantics

Review existing expert search
systems/approaches

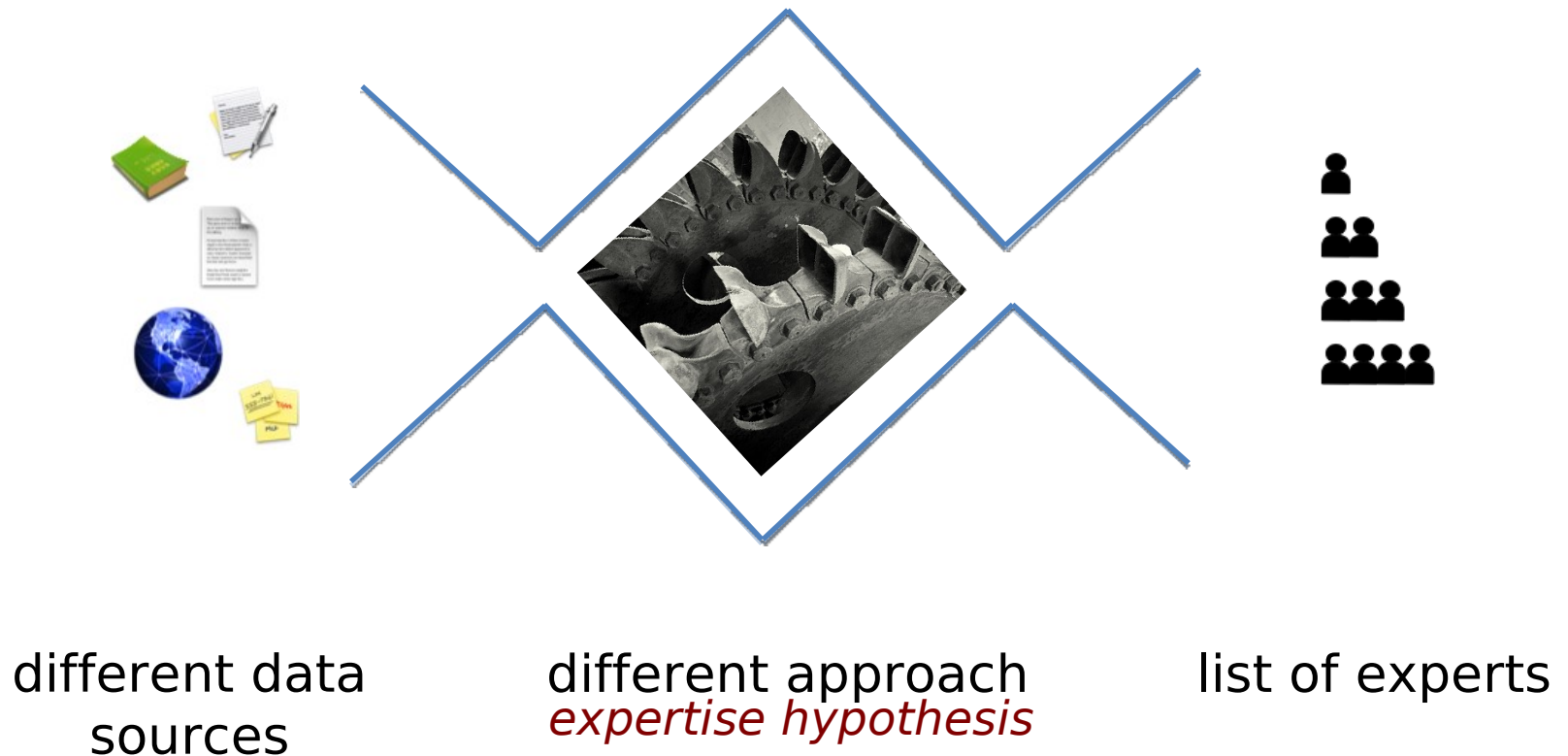
Extract and analyze expertise
hypothesis

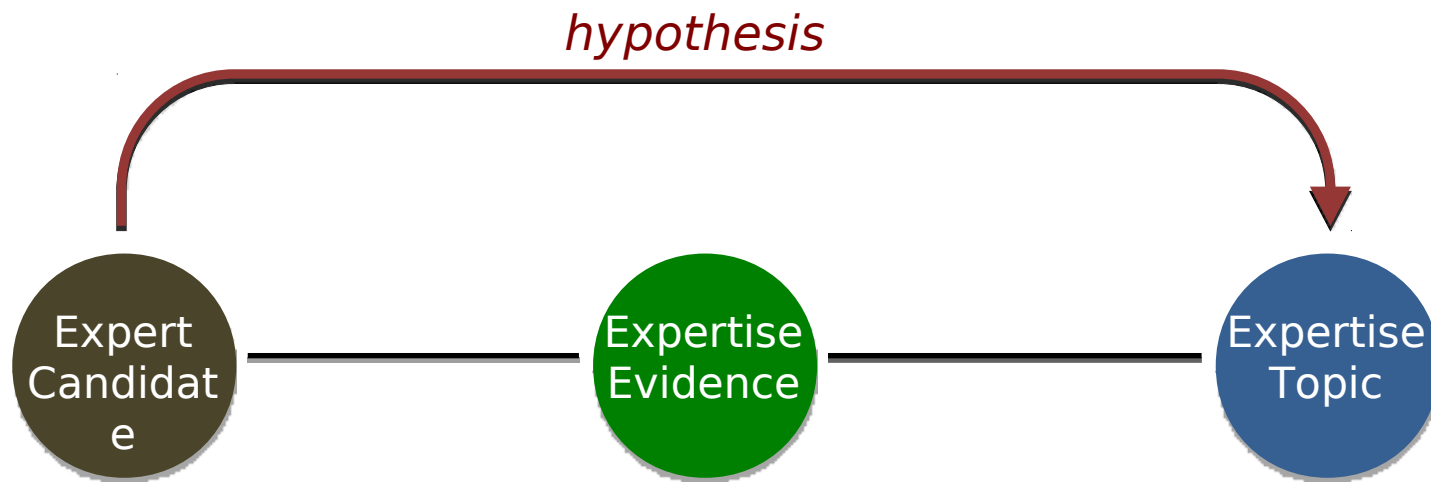
Test feasibility of expertise hypothesis on
LOD

Potentials and Pitfalls of LOD as
expertise evidence source



How do we search for experts in general?





Expertise Hypothesis

If the user	wrote a paper saved a bookmark saved a bookmark before the others was retweeted	on TopicX	then he/she is an expert then he/she is a better ranked expert	on TopicX
-------------	---	-----------	--	-----------



Content

Blogs, Publications,
Bookmarks, Wikipedia
Articles, ...

Activity

Attending professional
events, Roles on events,
Experience, Projects,...

Reputation

Social Connectedness,
User's popularity,
Quotes...

If a user wrote a scientific publication on topic X then he is an expert on topic X. If a user wrote a Wikipedia page on topic X then he is an expert on topic X. If a user edited or revised a document about topic X on a collaborative shared online workspace, then he might be expert on topic X. If a user blogs a lot about topic X, then he might be an expert for topic X. If a user has lower entropy of interests, where topic X is a primary interest, then he is a better expert on topic X. If a user has a lot of e-mails on topic X then he is an expert on topic X. If the user has resources/documents on topic X then he is an expert on topic X. If a user has subscription to feeds on topic X, then he is an expert in topic X. If a user participates in a Q&A community on a topic X then he is an expert on the topic X. If a user answers questions from experts then he might himself be an expert --> The more the user asking a question in a Q&A community is expert, the more significant is the expertise of the user giving the answer. If a user participates lots of email conversations about topic X then he might be an expert. If a user answers lots of questions about topic X then he is an expert on topic X. If the user discovers (and shares) "important/good" resources (i.e. resources which become later popular) on topic X, then he is an expert on topic X. If the user is among the first to find and share a good resource on topic X, then he is among the best experts on topic X. If the user participates in collaborative software development project then he might be an expert. If a user claims in his resume/CV that he is skilled in a topic then he might be expert.

Feasibility of expertise hypothesis on LOD?

If a user has obtained funded research grants in a certain (domain) field,

Test Cases



T1 Existence:

Does LOD contain data sets with the type of data needed for a certain hypothesis?



T2 Detail Level:

Are there relevant data in the concerned data sets?



T3 Interlinkage:

Are there any links to the topics of competence?

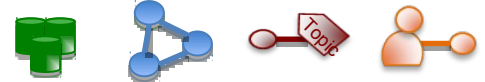


T4 Interlinkage:

Are there any links to a user's identities/accounts?

Test Results : Content

hypothesis related to content created by user



H1: If a user wrote a scientific publication on topic X than he might be an expert on topic X	+	+	+-	+
H2: If a user wrote a Wikipedia page on topic X than he might be an expert on topic X.	+	+	+	-
H3: If a user blogs a lot about topic X, then he might be an expert for topic X	+	+	+-	+-



T1: Does LOD contain data sets with the type of data needed for a certain hypothesis?



T2: Are there relevant data in the concerned data sets?



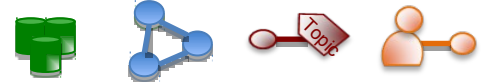
T3: Are there any links to the topics of competence?



T4: Are there any links to the user data sources?

Test Results: Online Activities

hypothesis related to users' online activities



H4: If a user answers questions (on topic X) from experts on topic X then he might himself be an expert on topic X	+	-	-	-
H5: If a user is among the first to discover (and share) "important/good" resources (i.e. resources which become later popular) on topic X, then he might be an expert on topic X.	+	-	+	-
H6: If a user participates in collaborative software development project then he might be an expert in the programming language that is used in the project.	+	+	+-	+-



T1: Does LOD contain data sets with the type of data needed for a certain hypothesis?



T2: Are there relevant data in the concerned data sets?



T3: Are there any links to the topics of competence?



T4: Are there any links to the user data sources?

Test Results: Offline Activities

hypothesis related to users' offline activities & achievements

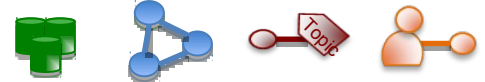


H7 If a user claims in his resume/CV that he is skilled in a topic X than he might be expert in topic X.	-	-	-	-
H8: If a user has obtained funded research grants in a certain (domain) field, then he might be an expert in that field.	+	+	-	+
H9: If a user has a certain position in company then he might be an expert on the topic related to his position.	+	-	-	+-
H10: If a user supervises/teaches someone then he might be an expert on the topic he/she teaches.	-	-	-	-
H11: If a user has several years of experience with working on something related to topic X then he might be an expert in topic X.	-	-	-	-
H12: If a user is a member of the organization committee of a professional event, then he might be expert on the topic of the event.	+	+	-	+
H13: If a user is giving a keynote or invited talk at a professional event, then he can be considered an expert in the domain topic of the event.	+	+	-	+
H14: If a user is a chair of a session within a professional event, then he can be considered an expert in the topic of the session (and by generalization, also an expert in the domain topic of the event).	+	+	-	+
H15: If a user is presenting within a session of a professional event, then he can be considered an expert in the topic his presentation is	+	+	-	+

T1: Does LOD contain data sets with the type of data needed for a certain hypothesis? T2: Are there relevant data in the concerned data sets? T3: Are there any links to the topics of competence? T4: Are there any links to the user data sources?

Test Results: Reputation

hypothesis related to users' reputation



H17: If a user's blog about a topic X gets lost of comments, then he might be an expert for topic X.	+	+	+-	+-
H18: If a user has higher social connectedness with an expert in topic X, then he is considered to be a better expert in topic X	+	+	+-	+-
H17: If a user's blog about a topic X gets lost of comments, then he might be an expert for topic X.	+	+	+-	+-



T1: Does LOD contain data sets with the type of data needed for a certain hypothesis?



T2: Are there relevant data in the concerned data sets?



T3: Are there any links to the topics of competence?



T4: Are there any links to the user data sources?

Summary of Results

- ♦ Content-based Hypothesis:
 - ♦ e.g., DBLP, SW Conference, SIOC, Faviki
- ♦ Reputation-based Hypothesis:
 - ♦ e.g., FOAF, SIOC
- ♦ Activity-based Hypothesis
 - ♦ e.g., SW Conference, DOAP Store
- ♦ Problems
 - ♦ Lack of details
 - ♦ Lack of Interlinkage (topics and users)

Potential Benefits

- ♦ Cross-Platform:
 - ♦ Complex hypothesis across different data sources
- ♦ Reuseable:
 - ♦ Decouple hypothesis and data sources
- ♦ Extensible and Flexible:
 - ♦ Discover new data sources for given hypothesis

Conclusion

- ♦ More data sources (especially about activities)
- ♦ More details (especially context information)
- ♦ Data descriptions for automatic data source selection
- ♦ Interlinks (user identities and topics)



Not just a critique, but a call for action!



Thank you for your attention.
milan.stankovic@hypios.com