ABSTRACT
So far, Web search has been a playground for few giants. However, while traditional search engines are superb in their ability of extracting the Web pages that most closely match with user’s keywords, they fail in going beyond such simple paradigm. On the other side, an increasing number of data sets is becoming available on the Web as (semi) structured data instead of user-consumable pages. Web search has huge potentials for improvement thanks to the high quality of these data sources, but this can be achieved only by designing new search applications that federate those sources. To tackle the long tail of user requirements and tastes, the need arises for new ways of thinking and designing search applications: application providers (and perhaps even end users) will need to build their own, customized search experiences, by combining search services available on the Web at the purpose of solving specific search needs. Individual and collective social experience will be more and more influencing search results. In our work we investigate the social, economic, and behavioral trends that push towards a completely different interpretation of the search task on the Web. We discuss how the technology and the micro-economic models must change to face these challenges.

Categories and Subject Descriptors
H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – search process, query formulation

General Terms
Design, Economics, Human Factors, Languages, Theory.

Keywords
Search Engine, Linked Data, Social Computation, Orchestration.

1. INTRODUCTION
So far, Web search has been a playground for few giants – such as Google, Yahoo! and Bing – that relegate the other players to market niches. However, while traditional search engines are superb in their ability of extracting the Web pages that most closely match with user’s keywords, they fail in going beyond such simple paradigm. Traditional approaches to search are reaching their limits in terms of effectiveness and utility for the end users. This is mainly due to: the change of perspective of the users, searching the Web for objects/concepts and not for pages; the increased complexity of user searches, typically oriented to solving complex tasks, such as buying houses or trip planning; the increasing availability of structured data sources, typically with open access, which require completely different treatment with respect to web pages; and the growing size of “search spamming”.

In our work we investigate the social, economic, and behavioral trends that push towards a completely different interpretation of the search task on the Web. Our findings are based on some evidences [3, 4], as well as some visions extrapolating them. We discuss how the technology and the micro-economic models must change to face these challenges. The need arises for new ways of thinking and designing search applications: application providers (and perhaps even end users) will need to build their own, customized search experiences, by combining search services available on the Web at the purpose of solving specific search needs. Individual and collective social experience will be more and more influencing search results. The problem is extremely challenging because it involves expertise from very diverse disciplines and will have a huge impact on how end users will perceive and access information on the Web in the medium-long term.

2. TECHNOLOGY AND DATA PROVISIONING EVOLUTION
An increasing number of data sets is becoming available on the Web as (semi) structured data instead of user-consumable pages. Linked Data plays a central role in this, thanks to initiatives such as W3C Linked Open Data (LOD) community project, which are fostering LD best practice adoption. LD and other open or proprietary data are made available through Web APIs and/or search-specific languages (e.g., see Google Places API or the Yahoo Query Language (YQL) framework, or languages like SQL, SPARQL, or XQuery). Methods, models and tools must be devised for efficiently designing and deploying new search services and data access APIs. Web search has huge potentials for improving the quality of search results thanks to the high quality of these data sources, but this can be achieved only by designing new search applications that federate and compose search services in the proper way, through complex and optimized Web query strategies [2].

3. APPLICATION DEVELOPMENT
Application developers could act as brokers of new search applications built by assembling search services. Along with powerful, generic, and world-wide applications (e.g., general purpose search engines or other generic utilities, such as geo-
localization services), we envision a growing market of specialized, localized, and sophisticated search applications, addressing the long tail of search needs (e.g., the “gourmet suggestions” about slow-food offers in given geographic regions). In this vision, large communities of service providers and brokers (e.g., like ProgrammableWeb.com and Mashape.com for mashups) could be empowered, with appropriate regulations concerning access rights and revenue sharing. Building support design environment and tools for executing orchestrations that compose search services is a challenging issue too.

Service orchestration is driven by both data and control dependencies, and strategy optimization must consider rank aggregation, quality of service (based e.g., on service statistics), optimal result paging, and so on [1]. To be concrete, think about developing a real estate search system in a given region, by using both global and local resources. The high-level orchestrations of search services will consist in the composition of services providing real estate offers (e.g., Trulia.com or Zillow.com), services on public utilities (e.g., Metro.com for public transportation) and services on statistics (e.g., WalkScore.com describing if the neighborhood is walking-friendly, in terms of services, shops, and restaurants at walking distance; or LocalCensus.com, extracting demographics and crime rate information). Notice that this is not a simple mashup of services, as it includes, e.g., complex optimization strategies and result composition logics [2], as described in section 4.

4. EFFICIENT ORCHESTRATION
When data sources must be joined, the join operation must take into account ranking; join can either be based on exact methods, according to the rank-join theory, or on approximate methods, that favor the speed of result production [7]. Data integration strategies should aim at obtaining a suitable number of results (e.g., neither too few, nor too many). Normally, a computation should not be set up so as to exhaust a searchable data source, as the user is rarely interested to inspect all of them. The ultimate controller is the user, who sees service results or their compositions and can halt their production.

Fig. 1: Search Service integration example, high level view.

In particular, if you refer to the example mentioned above, one can decide an orchestration as shown in Figure 1, where the Trulia and WalkScore services are invoked in parallel to extract housing offers and the walkability index of the various districts, thus extracting at most 30 good results by using at most 10 service calls. Then, results are piped to the Metro and LocalCensus services, which extract demographics and crime rate information for each solution. Three global thresholds are set on the number of results, on the time limit and on the number of service calls (30 results, 5 seconds, 50 calls respectively). Execution halts if any threshold is met.

From a system-oriented point of view, search-oriented engines for data integration should be deployed in “cloud computing” environments, to enable flexibility and scalability for application developers. In this way, the complexity of engine design and optimization would be totally outsourced and become transparent to the developer. In essence, the developer should not only resort to external data sources, but also to external data integration systems.

5. MICRO-ECONOMIC MODELS
The current business model of search is largely based on online advertisement. In today’s Internet advertising industry, the so-called search format is the most relevant revenue generating context: advertisers pay search engine companies to list their links (aka, sponsored links) in response to specific search word or phrases, with a bidding mechanism. Also cross-dependencies between search and sponsored results must be considered [5].

Given the envisioned change in the search applications, the model of profits based upon accountability of the click-through and/or of actual commercial transactions (e.g., online purchases) can be redefined. Multi-domain queries may offer an important dimension for bidding in auction mechanisms, by associating bids to keywords only when specific domain combinations are present in a solution (e.g., when a movie appears together with a DVD rental instead of a theatre). The new search applications can also act as brokers for sponsored links by merging different lists in the attempt to rank in high position the links that are the most appropriate in combination with others, based upon advanced models of click probability.

6. USER BEHAVIOURAL ANALYSIS
User behavior studies largely investigated the search activities. Of main interest is the exploration of information, performed by humans for getting closer and closer to the information they look for. Exploratory search [9] studies users starting from uncertain information needs and progressively discovering both needs and information, with a mix of look-up, browsing, analysis and exploration activities. A variety of different tools now exist for information exploration: dynamic faceted taxonomies [6], topic exploration (e.g., Kosmix topical engine [8]), community feedback and social wisdom, linked data exploration applications, and so on. Exploratory search is particularly relevant with the increase of complexity of search interaction, as the exploratory steps can be used to partition the complexity of query formulation. The appropriate ways for mixing of static orchestration and flexible exploration should be studied, both from the software engineering and the user interface design perspective. A set of different visualization and resource navigation perspectives should be put in place to grant the best navigational experience to the user. Figure 2 shows a map visualization widget that displays complex items (e.g., composed by a metro station and some housing options), together with a global ranking of the items (barchart on the left) and a list of items with the respective scores at the bottom, from which selection and subsequent exploration steps can be performed.
7. SOCIAL COMPUTATION, SOCIAL BEHAVIOUR AND IMPACT ON SEARCH

Involving humans in search is at its infancy, but it may completely reshape the search paradigm. Based upon the fact that the human brain is incredibly powerful at solving complex tasks, several systems (e.g., Yahoo Answers, Amazon Mechanical Turk) have explored the possibility of involving users in solving complex search problems. At the same time, social networks have emerged as one of the most relevant and trusted source of recommendation data, e.g. about resources such as hotels, restaurants, movies, even professors in course delivery. Integrating individual and collective human input into search computation is therefore the main challenge for the future of search. Finally, also the real-life social scenarios where search plays a role today must be taken into account. As an example, the “just Google it” attitude to address an information search problem is quickly changing to a more diverse behavior, where users choose a different search application (e.g., mobile search, augmented reality, social recommendation system,…) based on their current context and information need.

8. ACKNOWLEDGMENTS

This research is part of the Search Computing (SeCo) project, funded by ERC, under the 2008 Call for "IDEAS Advanced Grants" (http://www.search-computing.org). We wish to thank all the contributors to the project.

9. REFERENCES


