Position Statement - Leonid Libkin (University of Edinburgh)

My main research area is data management, in particular query
languages, incomplete information, database design, data integration
and exchange, and non-relational data, including graph
databases. Relevant to graph databases, I investigated how to add path
comparisons to query languages, how to combine data and topology
querying over graphs, and how to use XPath ideas for querying
graphs. More recently, I led the Edinburgh part of the Neo4j/Edinburgh
team that provided a formal semantics of the Cypher language. We
propose that three members of the Cypher semantics group be present at
the workshop: Paolo Guagliardo, my colleague from Edinburgh, and
Victor Marsault, formerly of Edinburgh, currently in Paris.

The topic we would like to lead the discussion on is formal semantics
of query languages. We shall describe our experience in formalizing
Cypher, as well as the core of SQL, and reflect on how it can be used
to provide a formal model and semantics of the forthcoming GQL. For
languages like SQL, the Standard came first, and formal semantics
appeared much later. The Standard written in natural language may well
lead to ambiguities, which are reflected in well-known discrepancies
between SQL implementations by major DBMSs. In the case of Cypher the
semantics appeared much earlier; it led to a formal data model as well
as precisely specified meaning of all language constructs. In addition
to it establishing the yardstick against which other implementations
can be measured, it also led to rethinking and reimplementing several
features of Cypher. Furthermore, the semantics of Cypher comes with a
tool that implements it, and thus can be used to evaluate the
compliance of other products that use Cypher with the description of
the language, by means of a suite of test queries that can be run and
then compared with the correct results provided by the semantics of
the language.

We shall argue that the semantics of the language should be developed
in parallel with the development of the new Standard, not as an
afterthought. It will not replace the Standard but should, in our
view, become a part of it, together with the natural language
description. Similarly to the experience we had with Cypher, it will
guide language development, help avoid ambiguities, lead to the
discovery of tricky corner cases, and provide a tool for checking
compliance of different implementations with the Standard.

--
The University of Edinburgh is a charitable body, registered in
Scotland, with registration number SC005336.