Querying Past and Future in Web Applications

Daniel Deutch Tova Milo

Tel-Aviv University



Outline

- Introduction & Motivation
- Querying Future
 [VLDB'06, ICDE'09, ICDT'09, VLDB'09]
- Querying Past
- [VLDB'07, VLDB'08, ICDT'09, VLDB'09]
- Related and Future work

Outline

Introduction & Motivation

- Querying Future
 [VLDB'06, ICDE'09, ICDT'09, VLDB'09]
- Querying Past
- [VLDB'07, VLDB'08, ICDT'09, VLDB'09]
- Related and Future work

Querying Web applications

- Web-based Business Processes (BPs) are very popular
- Querying the past
 - past executions of a given application for improving business logic, optimization, personalized ads,...
 - past design patterns of applications to exploit when building a new application
- Querying (possible) future executions
 - the above + verification

Example Web Application

Cars/Rail Vacation Packages Cruises Last Minute Packages

😞 Lufthansa

search select review reserve confirm

Sort by Travel Ti

Total travel time: 1 hrs 15 mins

Hotwire Save up to 35% on flights O

, Aer Lingus

0 nonstops

\$574

1 flight

Activities

C KLM Packages from \$899

Help with this page

Customer Care | My Stu

Book Online or Call: 888-Travelocity

Flight + Hotel

\$756 Select

* travelocity	Your Flight to London, Great Britain (LON) Departing: Tue, Apr 22 - Returning: Tue, Apr 29 1 Aduit Change Your Search Your Search Depart Tue, Apr 22 Flights + 7 Nights Hotel Save with TotaTrip ^{6M} from \$209
Home Vacation Packages Flights Hotels Travel Info Center Flight Status Destination Guides Training	Information about your flight results Flights departing close to your preferred time are highlighted below. See how to gort be Prices shown are for <u>e-tickets</u> . Paper tickets require <u>additional fees</u> .
Packages Hotels Cars Flights	Nonstops Only \$209 \$215 0 nonstops 0 nonstops
Book Flight & Hotel	All 37 Flights \$209 \$215 \$382 \$4 displayed below 11 flights 15 flights 4 flights 4 flights
	Home Flights Hotels Cars/Rail Vacation Packages Cruises Last Min
From: To:	Your 8 Day, 7 Night Trip
london paris	Your 8 Day, 7 Night Trip Search select revi London Heathrow Airport (LHR) to Paris Charles de Gaulle Airport (CDG) Departing Tue, Apr 22 Returning Tue, Apr 29 1 Room, 7 Nights 1 Aduit London ge search London ge search Returning Tue, Apr 29 1 Aduit
Io: Iondon paris	Search select rev Your 8 Day, 7 Night Trip London Heathrow Airport (LHR) to Paris Charles de Gaulle Airport (CDG) Departing Tuo, Apr 22, Returning Tue, Apr 29 1 Room, 7 Nights 1 Aduit Change search Flight - 7 Nights Hotel Flight - 7 Nights Hotel Flight - 7 Nights Hotel
Compare surrounding airports Exact dates +/- 1 to 3 days Flexible dates	Search Select rew Your 8 Day, 7 Night Trip Landon Heathrow Asport (LHR) to Peris Charles de Gaulle Airport (CDG) Departing Tue, Apr 22 - Returning Tue, Apr 29 1 Room, 7 Nights Select Your Outbound Flight with this Hotel Hotel - 1 Room, 7 Nights Tue, Apr 29 Check Nu Days Inn Aulhay-sous-Bois, Auhay sous Bois, Tue, Apr 29

Application Specification

- BP specifications (e.g. in BPEL) are compiled to running code
 => Queries over specification structure
- Modeled as nested DAGs
 Each DAG corresponds to a function\web-page
- Nodes model activities (activation and completion)
 Edges mark flow relation
 Atomic/compound activities
 Nesting models implementation relations \ links
 Guarding formulas (on external events) model choices
- Recursion is allowed

Introduction and Motivation Example Specification



Introduction and Motivation Example Execution Flow



Sources of Uncertainty

Partial Tracing, due to lack of storage, confidentiality, etc.

- External Effects, e.g. user choices, server response time
 - The effect of external events is modeled by logical formulas, guarding implementations
 - At run-time, formulas' truth values determines the chosen implementations

Thus the **past** is uncertain, and the **future** unknown



Queries on possible EX-flows

- Can a user reserve a flight without giving her credit card details?
- What is the **probability** that this would happen?
- How can this happen / what are the most common scenarios?
- What is the common behavior of users that search for an Air France flight+hotel deal but quit without making a reservation?

Focus on the search-related sub-flow.

What is the best way to get a cheap flight+hotel deal, or one that maximizes the FF millage/points ?









Query language

- Execution patterns
- Intuitive, similar in structure to execution flows
- Seek for occurrences (homomorphism) of the pattern within execution flow DAGs
- May contain transitive nodes and edges
- May contain a projection part

Example Query



Weighting flows

- A weighted model
 - CWeight choice Weight (product cost, shipping time, likelihood)
 - Aggr (sum, multiplication)
 - fWeight flow Weight (total cost, shipping time,...)

- Varying sensitivity to flow history (cWeight)
- Varying level of monotonicity (fWeight)

Our Example Specification



Querying Future Example (Weight = Likelihood)

- We distinct three classes of probability distributions, according to the level of dependency of cWeight.
- History independence (markovian): No dependencies between formulas.
- Bounded-history: Dependency in(at most) B last formula values.
- Unbounded-history

Sample of Results (top-k selection)

Complexity of query evaluation depends on cWeight

- PTIME in the spec., exponential in history bound and query.
- NP hard in both.

Querying Future

- Undecidable for unbounded history
- (Instance) Optimality of our algo depends on fWeight
 - Strongly monotone: Optimal
 - Semi-strongly monotone: Instance optimal

No optimal exists

Weakly monotone:

Not (instance) optimal No instance optimal exists

Sample of Results (Boolean)

- Harder:
 - Need to sum up probability of (a possibly infinite number of) qualifying flows
- Computing exact probability is impossible
- Approximation is possible
 - Technique: representing probabilities via set of linear equations
 - EXPTIME in general (NP Hard)
 - PTIME for non-recursive apps

Querying Future Sample of Results (top-k projection)

Source of difficulty:

- need to consider a possibly infinite # of answers,
- and sum up probabilities of their possibly infinite # of origins

Technique:

- "Small world" theorem allows to consider only a bounded number of answers,
- then use (Boolean queries) Oracle

EXPTIME even for non recursive apps (NP Hard)

ShopIT – Shopping assistanT



ShopIT – Shopping assistanT



ShopIT – Shopping assistanT





Types of Partial Traces

- Partial Tracing, due to lack of storage, confidentiality,...
 - **Naïve** tracing records all activities accurately
 - Semi-Naïve tracing contains only partial information on the names of some activities
 - **Selective** tracing may omit some activities occurrences
- Tracing systems (called types) are represented by a renaming function and a deletion set

Example BP



Naïve Traces



Semi-Naïve Traces



Selective Traces



Querying Past Lets talk about (top-k) queries

- Given a partial trace, what is its most likely origin?
- Or, more generally, given a pattern (query) of partial traces, what are the most likely origins of partial traces of this pattern ?

Good news:

Many of the query evaluation algorithms extend to this context (even without knowing the tracing system...)



Related & Future Work (Small subset of) Related work

- (Probabilistic) Recursive State Machines with temporal logic as query language
- Probabilistic Relational DBs
- Probabilistic XML
- Graph grammars with MSO (or FO) as query language
- BP and Web applications mining

Related & Future Work

Future work

Practical applications:

- Web-sites design
- On-line advertisements
- Improved business logic

Enriched Query Language

- Joins
- Data values

Optimization

Inference of specs/probability distributions

