Evaluating Preference-based Search Tools: A Tale of Two Approaches

Paolo Viappiani,[1] Boi Faltings,[1] and Pearl Pu[2]
[1]Artificial Intelligence Lab (LIA)
[2]Human Computer Interaction Group (HCI)
Ecole Polytechnique Federale Lausanne (EPFL)

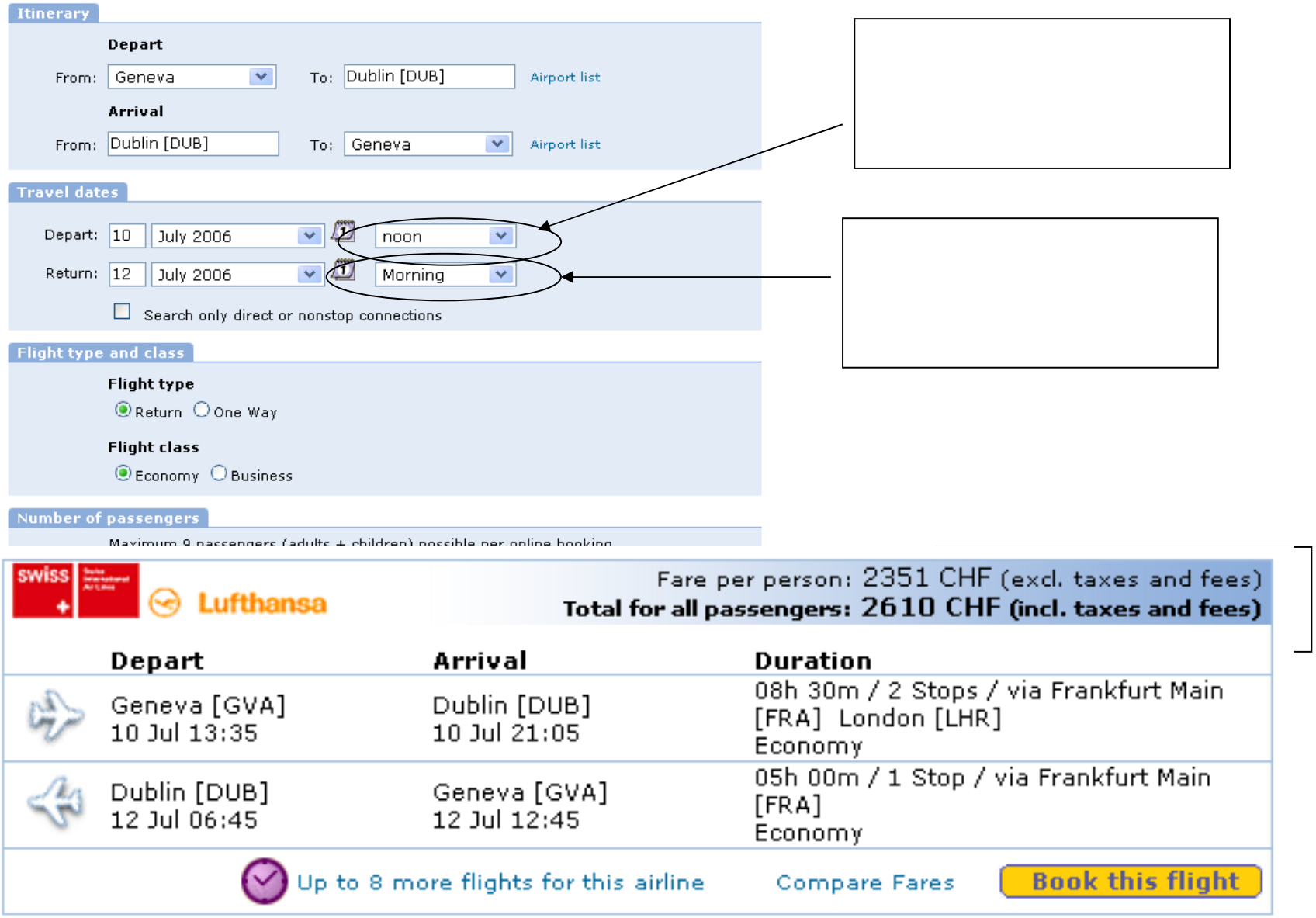
AAAI 2006, **The Twenty-First National Conference on Artificial Intelligence, July, 16-20, Boston, MA, USA**

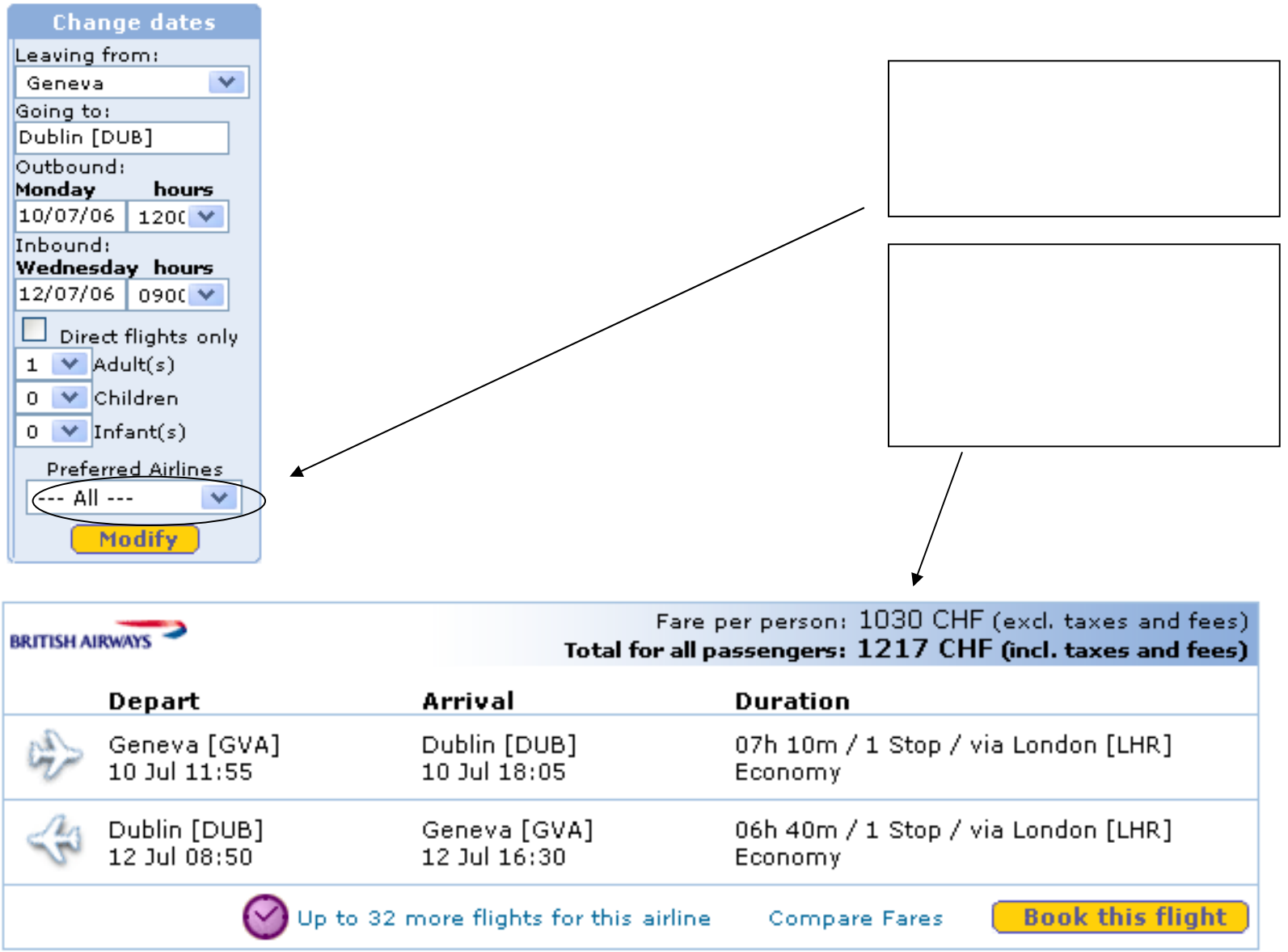
Preference-based Search

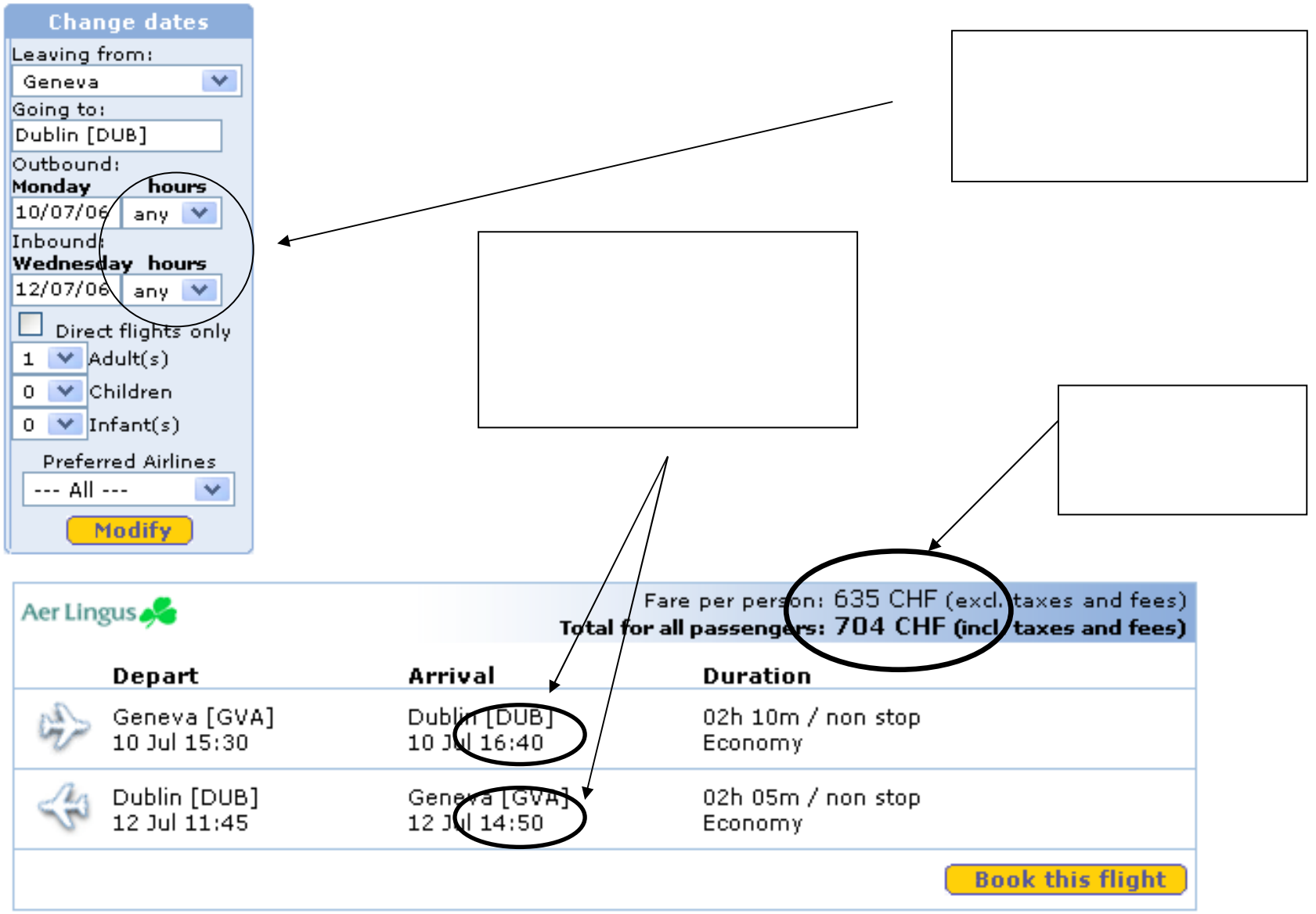
- People often use the **www** to search for their most preferred item
  - Computers, cameras, apartments, flights
  - Structured items can be searched in a database
- **Crucial**: an accurate model of users' preferences
  - Classic procedures for utility elicitation (Keeney) require too much effort
  - Most common approach is to ask the user to fill in a form

Example

- Actual scenario with travel website (July 5th, 2006)
- User wants to travel from Geneva to Dublin
- Return flight
- Preferences
  - Outbound flight, arrive by 5pm
  - Inbound flight, arrive by 3pm
  - (Cheapest)


Itinerary
Depart
From: Geneva
To: Dublin [DUB]
Airport list
Arrival
From: Dublin [DUB]
To: Geneva
Airport list
Travel dates
Depart: 10 July 2006 noon
Return: 12 July 2006 Morning
Search only direct or nonstop connections
Flight type and class
Flight type
Return
One Way
Flight class
Economy
Business
Number of passengers
Maximum 9 passengers (adults + children) possible per online booking
SWISS
Lufthansa
Fare per person: 2351 CHF (excl. taxes and fees)
Total for all passengers: 2610 CHF (incl. taxes and fees)
Depart
Arrival
Duration
Geneva [GVA]
10 Jul 13:35
Dublin [DUB]
10 Jul 21:05
08h 30m / 2 Stops / via Frankfurt Main [FRA] London [LHR]
Economy
Dublin [DUB]
12 Jul 06:45
Geneva [GVA]
12 Jul 12:45
05h 00m / 1 Stop / via Frankfurt Main [FRA]
Economy
Up to 8 more flights for this airline
Compare Fares
Book this flight


Change dates
Leaving from:
Geneva
Going to:
Dublin [DUB]
Outbound:
Monday hours
10/07/06 120(
Inbound:
Wednesday hours
12/07/06 090(
Direct flights only
1 Adult(s)
0 Children
0 Infant(s)
Preferred Airlines
--- All ---
Modify
BRITISH AIRWAYS
Fare per person: 1030 CHF (excl. taxes and fees)
Total for all passengers: 1217 CHF (incl. taxes and fees)
Depart
Arrival
Duration
Geneva [GVA]
10 Jul 11:55
Dublin [DUB]
10 Jul 18:05
07h 10m / 1 Stop / via London [LHR]
Economy
Dublin [DUB]
12 Jul 08:50
Geneva [GVA]
12 Jul 16:30
06h 40m / 1 Stop / via London [LHR]
Economy
Up to 32 more flights for this airline
Compare Fares
Book this flight


Change dates
Leaving from:
Geneva
Going to:
Dublin [DUB]
Outbound:
Monday
hours
10/07/06
any
Inbound:
Wednesday
hours
12/07/06
any
Direct flights only
1
Adult(s)
0
Children
0
Infant(s)
Preferred Airlines
--- All ---
Modify
Aer Lingus
Fare per person: 635 CHF (excl. taxes and fees)
Total for all passengers: 704 CHF (incl. taxes and fees)
Depart
Arrival
Duration
Geneva [GVA]
10 Jul 15:30
Dublin [DUB]
10 Jul 16:40
02h 10m / non stop
Economy
Dublin [DUB]
12 Jul 11:45
Geneva [GVA]
12 Jul 14:50
02h 05m / non stop
Economy
Book this flight

Form-filling is not effective

- Incorrect means objectives: formulate the real goal by a “substitute” goal believed to lead to desired outcome
- Users often state more preferences than necessary when prompted
- ***The preference model may be complete, but not accurate***

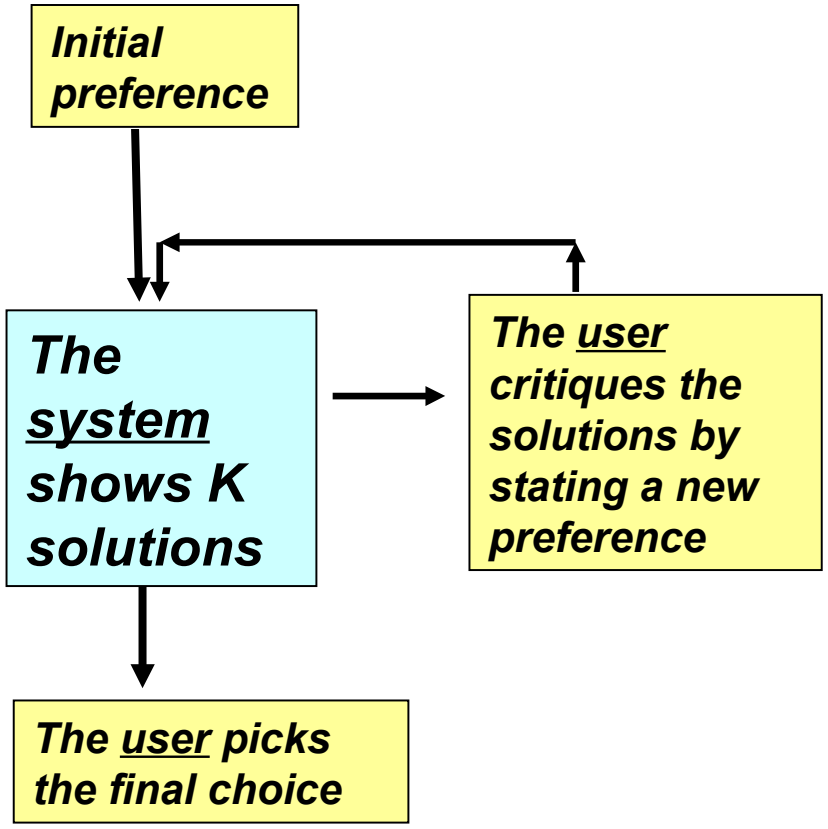
Alternative: preference construction

- Users' preferences are often constructed when considering specific examples
  - behavioral decision theory (Payne et al. '93; Slovic'95; Tversky '96)

- Collaborative filtering recommends items based users' rating on similar items
  - When users volunteer to rate items, more accurate recommendations are given *(McNee et al. '03)*

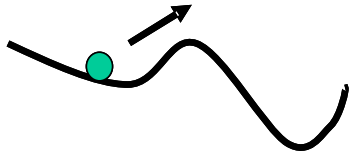
Allow users to self-initiate preference expression

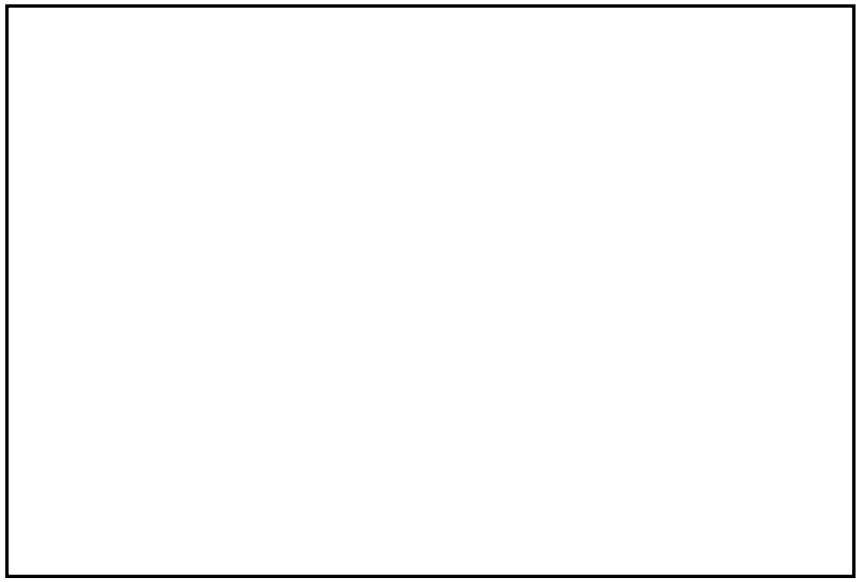
Example-based tools



- Several proposed systems:
  - Findme *(Burke et al. '97)*
  - Smartclient *(Pu&Faltings'00)*
  - Expertclerk *(Shimazu'01)*
- User expresses the preferences as *critiques* on displayed examples
- Feedback directs the next search cycle
- Users are more motivated to express preferences when self-initiated
  - *Suggestions*

The need for Suggestions





- Others have also recognized the need to help users consider potentially neglected attributes
- Show *extreme* examples *(Linden'97)*
- Show diverse examples *(*Smyth &McGinty'03, *McSherry'02)*
- Show suggestions based on the current preference model and possible extensions *(Pu et al. '06): model-based suggestion*
  - Optimally stimulate preference expression
  - Metaphor of Active Learning

Example Critiquing vs. Form Filling

Via user studies, we ask

- Do EC tools achieve better decision accuracy than traditional form-filling approaches?
- Are preferences more accurate when they were obtained from example critiquing vs. form-filling?

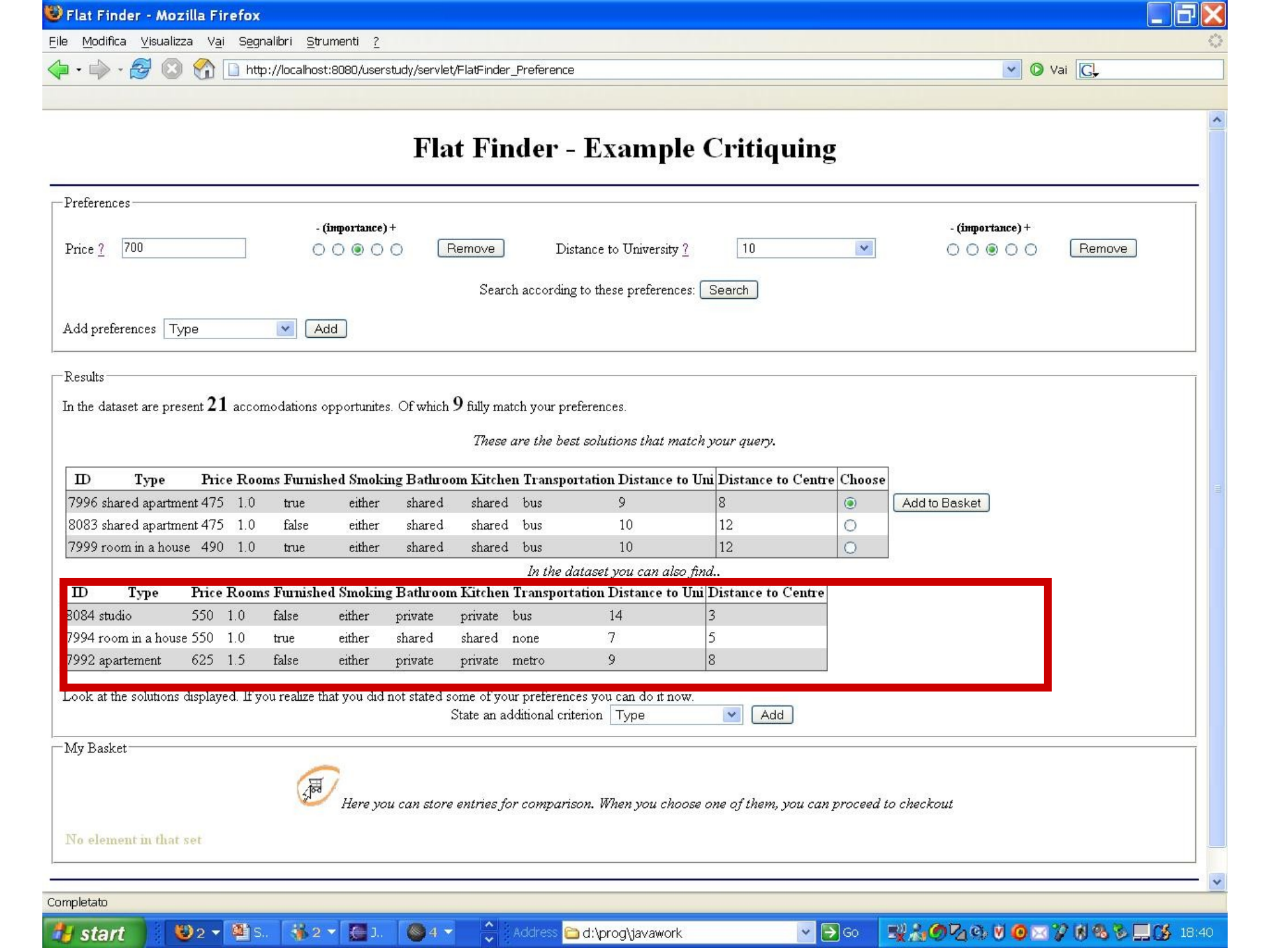
User Studies

- **60 users** searched their most preferred item from **180 items** in a database

- Measured variables
  - decision accuracy (Pu&Chen '05) : the percentage of times the system succeeded in finding users' most preferred item
  - user effort: the task time a user takes while using the tool to reach an option that she believes to be the target item

User Studies: Experiments

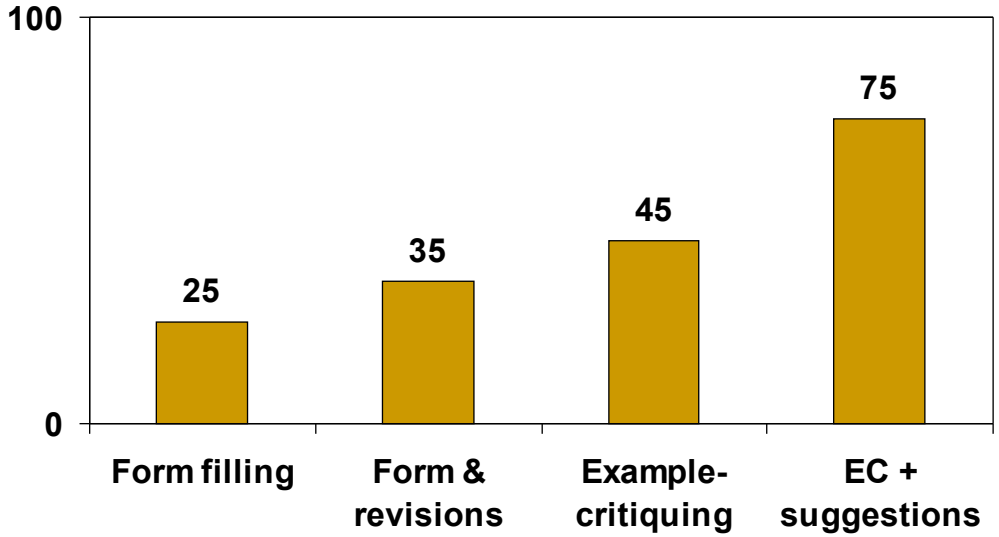
- Between-groups experiment (3 groups of 20 people)
  - **Form-filling interface:** user selects a preferred value or "don't care" choice on each attribute
  - **Example-critiquing interface:** user only states self-initiated preferences**;** views 6 best options
  - **Example-critiquing interface with suggestions:** user only states self-initiated preferences; views 3 best options and 3 suggestions
- Within-subject experiment (20 users)
  - **Form-filling interface**
  - **Example-critiquing interface**: showing 3 best options and 3 suggestions


Flat Finder - Mozilla Firefox
File Modifica Visualizza Vai Segnalibri Strumenti ?
http://localhost:8080/userstudy/servlet/FlatFinder_Preference
Vai
Flat Finder - Example Critiquing
Preferences
Price ? 700
- (importance) +
Remove
Distance to University ? 10
- (importance) +
Remove
Search according to these preferences: Search
Add preferences Type Add
Results
In the dataset are present 21 accomodations opportunites. Of which 9 fully match your preferences.
These are the best solutions that match your query.
ID Type Price Rooms Furnished Smoking Bathroom Kitchen Transportation Distance to Uni Distance to Centre Choose
7996 shared apartment 475 1.0 true either shared shared bus 9 8
8083 shared apartment 475 1.0 false either shared shared bus 10 12
7999 room in a house 490 1.0 true either shared shared bus 10 12
Add to Basket
In the dataset you can also find..
ID Type Price Rooms Furnished Smoking Bathroom Kitchen Transportation Distance to Uni Distance to Centre
8084 studio 550 1.0 false either private private bus 14 3
7994 room in a house 550 1.0 true either shared shared none 7 5
7992 apartement 625 1.5 false either private private metro 9 8
Look at the solutions displayed. If you realize that you did not stated some of your preferences you can do it now.
State an additional criterion Type Add
My Basket
Here you can store entries for comparison. When you choose one of them, you can proceed to checkout
No element in that set
Completato
start
Address d:\prog\javawork
Go
18:40

Between groups Experiment

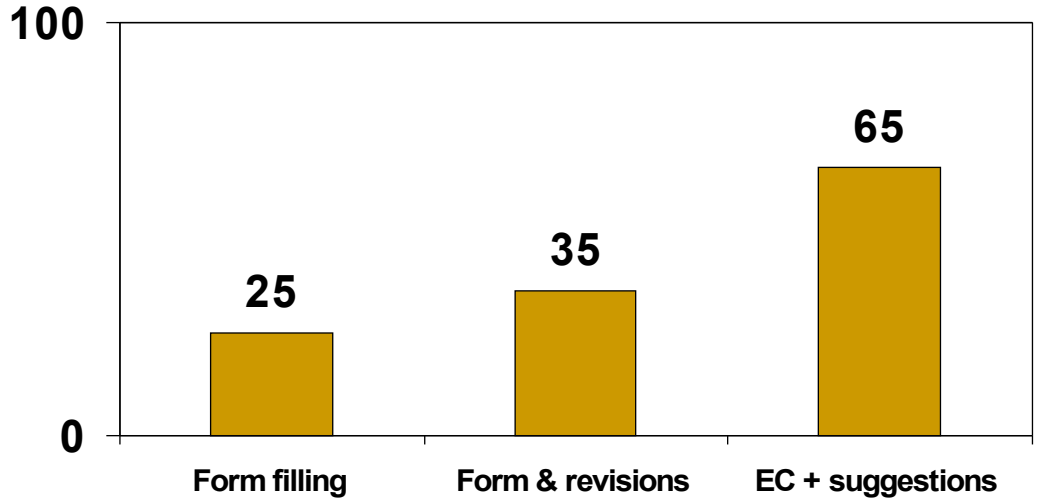
Accuracy increases with suggestions

- EC with suggestion
  - Better than form-filling ($p<0.01$)
  - Better than EC without suggestions ($p<0.02$)
- EC without suggestions
  - Better than Form filling (but $p>0.05$)



Time	2:45	5:30	8:09	7:39
Cycles	1.0	2.2	5.6	6.3

Within-subject Experiment



Time (avg)	2:45	5:30	6:00
Cycles (avg)	1.0	2.2	5.2

Comments on the results

- **Form-filling**: users state average of 7.5 preferences
  - Before having considered any of the available options
  - Even after revisions, preferences were not retracted

- **EC**: users begin with average of only 2.7 preferences, added average of 2.6 to reach 5.3
  - 50 % preferences were added during interaction
  - Results suggest that volunteered preferences are more accurate

- More preference revisions → higher decision accuracy (Pu&Chen ’05)
  - People who found their targets made more revisions
  - 6.9 as opposed to 4.5, statistically significant (p=0.0439)

Conclusions: a tale of two approaches

- Do not ask too many questions
  - Even though form filling interfaces are easier to implement
- Show attractive suggestions
  - User effort should be compatible with motivation for decision accuracy
  - Model-based suggestions effectively stimulate users to express accurate preferences
- User study validates the hypotheses

	Fare (a_1)	Arrival (a_2)	Airport (a_3)	Airline (a_4)
O1	250	14:00	INT	B
O2	300	9:00	INT	A
O3	350	17:30	CITY	B
O4	400	12:30	CITY	B
O5	550	18:30	CITY	B
O6	600	8:30	CITY	A

-
  - Arrive by 12:00
  - Leave from City airport

	Fare (a1)	Arrival (a2)	δ_2	Airport (a3)	δ_3	Airline (a4)	δ_4	P
O1	250	14:00	-	INT	-	B	-	
O2	300	9:00	*0.5*	INT	*0*	A	0.5	0.437
O3	350	17:30	*0.35*	CITY	*0.5*	B	0	0.381
O4	400	12:30	*0*	CITY	*0*	B	0	0
O5	550	18:30	*0.1*	CITY	*0*	B	0	0.05
O6	600	8:30	*0.05*	CITY	*0*	A	0	0.025

$$P(o) = 1 - \prod_{a_i \in A_u} (1 - P_{a_i} \delta_i(\theta))$$