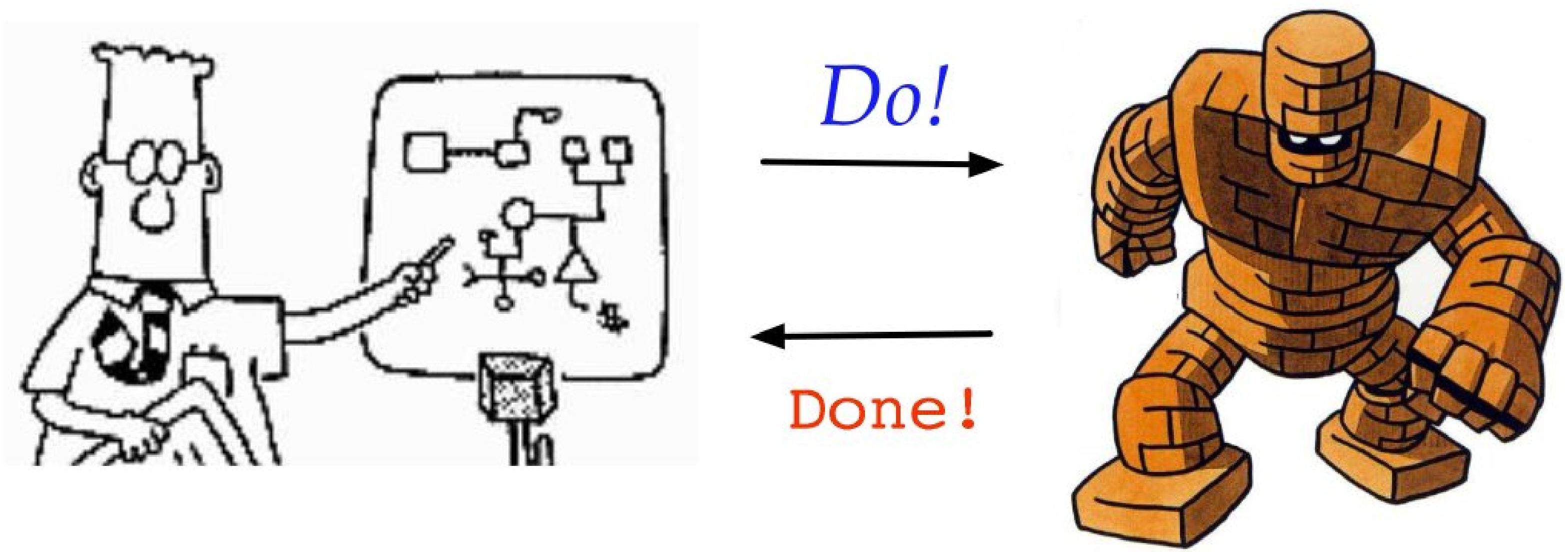
Eliciting GAI preference models with binary attributes aided by association rule mining

Sergio Queiroz - CIn/UFPE
srmq@cin.ufpe.br

Why preferences?

- Acting on behalf of a user...



Why preferences? (2)

- Simple goals aren't enough
  - They are rigid: do or die! (ex.: solve a puzzle)
- The world can be highly unknown
  - We can't know ahead of time if our ultimate goal is achievable
- So, what can we do?
  - We can go to the second best alternative
    - But what is “second best”?
    - And what if “second best” is infeasible?

How to do it?

- May be easy if there is a natural way to rank the alternatives
  - One objective with a natural order
    - Optimize cost, optimize quality
    - But, what about optimize both?
  - Adequate to very small sets of alternatives
    - Canarius Palace Hotel > Youth Hostel > A Bench at “Parque da Jaqueira”

But...

- Find the best vacation trip advertised in the web
  - Large space of alternatives
    - Lots of trip propositions advertised on the web
      - I don't want to view or compare all of them
  - Multiple objectives may be involved
    - Flight time, price, activities
  - Uncertainty about the feasible outcomes
    - Are there any offers for “one week in Tahiti” for under 200EUR out there?

Decisions in large spaces...

- Space of alternatives = Cartesian product
  - $X = X_1 \times X_2 \times \ldots \times X_n$ X_i finite domain of possible values
  - If each attribute domain has p values, size p^n
  - Represent preferences in extension needs a huge memory space
    - n = 10, p = 10 ➔ 10 GB; n = 10, p = 20 ➔ 10 TB
- Decision-making is very difficult (too many alternatives)

What we do...

- Take advantage of the structure of the preferences
  - Informally: User preferences have a lot of regularity (patterns) in terms of *X*
  - Formally: User preferences induce a significant amount of preferential independence over *X*
- Compact representations
- Benefits
  - Easier to elicit (construct) the model
  - Possibility to build efficient algorithms to exploit the model

Utility Functions

- Space of alternatives: $X = X_1 \times X_2 \times \ldots \times X_n$
- Appreciation (utility) of an alternative $x \in X$
  - $u : X \mapsto \mathbb{R}$
- Additive model $u(x) = \sum_{i=1}^{n} u_i(x_i)$
  - Simple and efficient: $p^n \rightarrow p \times n$
  - Independence between attributes

I prefer to drink red wine when I eat steak
but to drink white wine when I eat fish.

Utility Functions

- Space of alternatives: $X = X_1 \times X_2 \times ... \times X_n$
- Appreciation (utility) of an alternative $x \in X$
  - $u : X \mapsto \mathbb{R}$
- Additive model $u(x) = \sum_{i=1}^{n} u_i(x_i)$
  - Simple and efficient: $p^n \rightarrow p \times n$
  - Independence between attributes

I prefer to drink red wine when I eat steak
but to drink white wine when I eat fish.

Generalized Additive Independence (GAI)

- Definition (Fishburn, 70; Baccus and Groove 95)
  - $X = X_1 \times X_2 \times \ldots \times X_n$
  - $C_1, \ldots, C_k$ subsets of $N = \{1, \ldots, n\}$ such as

$$N = \bigcup_{i=1}^{k} C_i \quad , \quad X_{C_i} = \{X_j : j \in C_i\};\ u_i : X_{C_i} \rightarrow \mathbb{R}$$

A GAI utility function over X can be written in the form:

$$u(x_1, \ldots, x_n) = \sum_{i=1}^{k} u_i(x_{C_i})$$

Example:
$u(x_1, x_2, x_3, x_4) = u_1(x_1) + u_2(x_2, x_3) + u_3(x_3, x_4)$

Dependencies between attributes

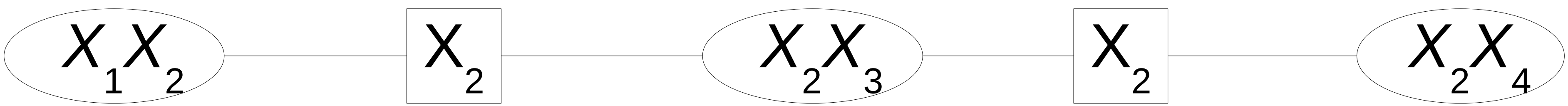
Non-disjoint factors

A Big problem:

- How to elicit a GAI model
  - Construct the model by asking questions about the decision maker preferences
    - They should be simple
    - They should be in a small number
- Inter-dependencies between attributes
  - Elicitation of each utility subterm separately is impracticable

Some relief: GAI Networks

- $u(x_1, x_2, x_3, x_4) = u_1(x_1, x_2) + u_2(x_2, x_3) + u_3(x_2, x_4)$

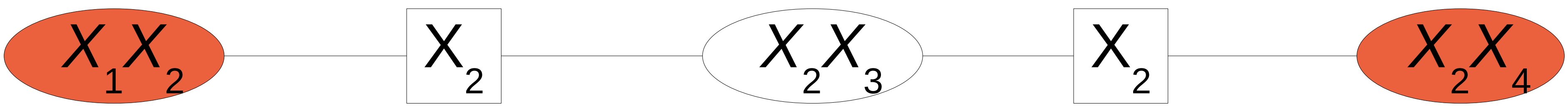


Ellipse = clique

Rectangle = separator

Some relief: GAI Networks

- $u(x_1, x_2, x_3, x_4) = u_1(x_1, x_2) + u_2(x_2, x_3) + u_3(x_2, x_4)$



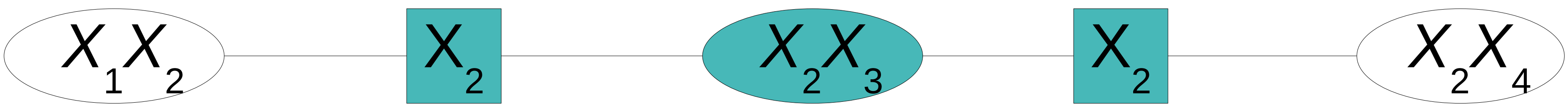
Ellipse = clique

Rectangle = separator

- Running intersection property:
  For any pair of cliques (C_1, C_2 with nonempty intersection S,

Some relief: GAI Networks

- $u(x_1, x_2, x_3, x_4) = u_1(x_1, x_2) + u_2(x_2, x_3) + u_3(x_2, x_4)$



Ellipse = clique

Rectangle = separator

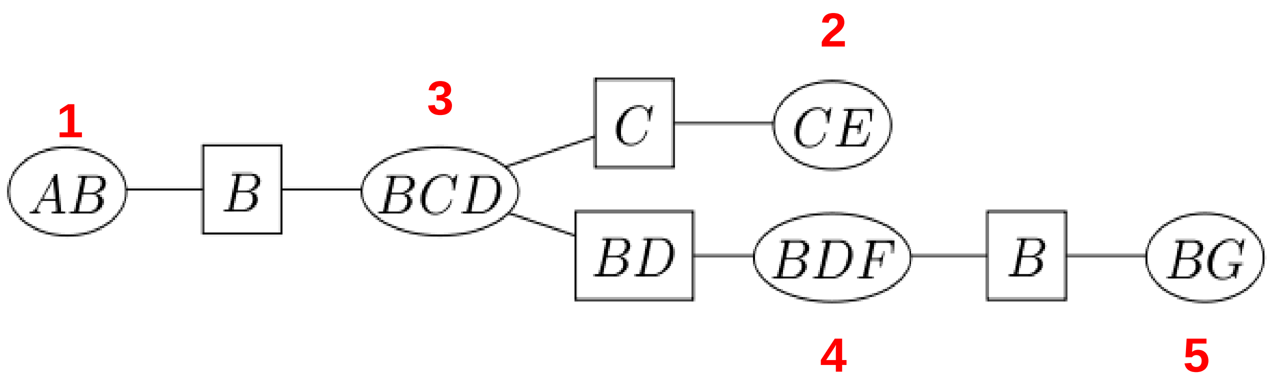
- Running intersection property:
  For any pair of cliques (C_1, C_2 with nonempty intersection S, S is a subset of every clique and separator on the path between C_1 and C_2
- Junction trees: constraint/Bayesian network literature

Why Junction Trees to Represent GAI functions

- Algorithmic efficacy in choice, ranking
  - Family of variable elimination algorithms
- Also allows elicitation with local questions

Elicitation under Certainty with GAI Networks - Notions

- Cliques ordered from exterior to interior ones



- Elicitation in three phases
  - Values given the instantiation of the separator
  - Intraclique
  - Intercliques

A problem that remains...

- What if the number of attributes is too high?
  - As even if they are all independent, I don't want to express my preferences over all of them
- Example
  - Visiting touristic sites in Paris
    - Binary attributes: touristic sites
      - ex.: Tour Eiffel, Musée du Louvre,...
      - (we found more than 200 of them)
    - Binary values: 0 (don't visit), 1 (visit)

Mining association rules to the rescue!

- Set of touristic sites: $I = \{i_1, i_2, ..., i_n\}$
- Set of trips to this destination: D
  - Each element of D (a trip) is a set of items $T \subseteq I$
- Associated mining problem
  - Set of literals: I
  - Set of transactions: T
  - Each transaction is the set of touristic sites in a trip
- Rules $X \Rightarrow Y$ (where $X \subset I$, $Y \subset I$ and $X \cap Y = \emptyset$)
  - People that visit items in X also visit items in Y

Coming to a non-linear 0-1 Knapsack problem

Maximize $u(x_1, \ldots, x_n) = \sum_{i=1}^{k} u_i(x_{C_i})$

under the constraint $\sum_{j=1}^{n} w_j x_j \leq c,$

GAI-decomposable function

- u_i is the utility function for the user preferences
- w_j is the time needed to visit item *j*
- x_j in {0, 1} (visit or not item *j*)
- *c* total time available in the trip

Coming to a non-linear 0-1 Knapsack problem

Maximize $u(x_1, \ldots, x_n) = \sum_{i=1}^{k} u_i(x_{C_i})$

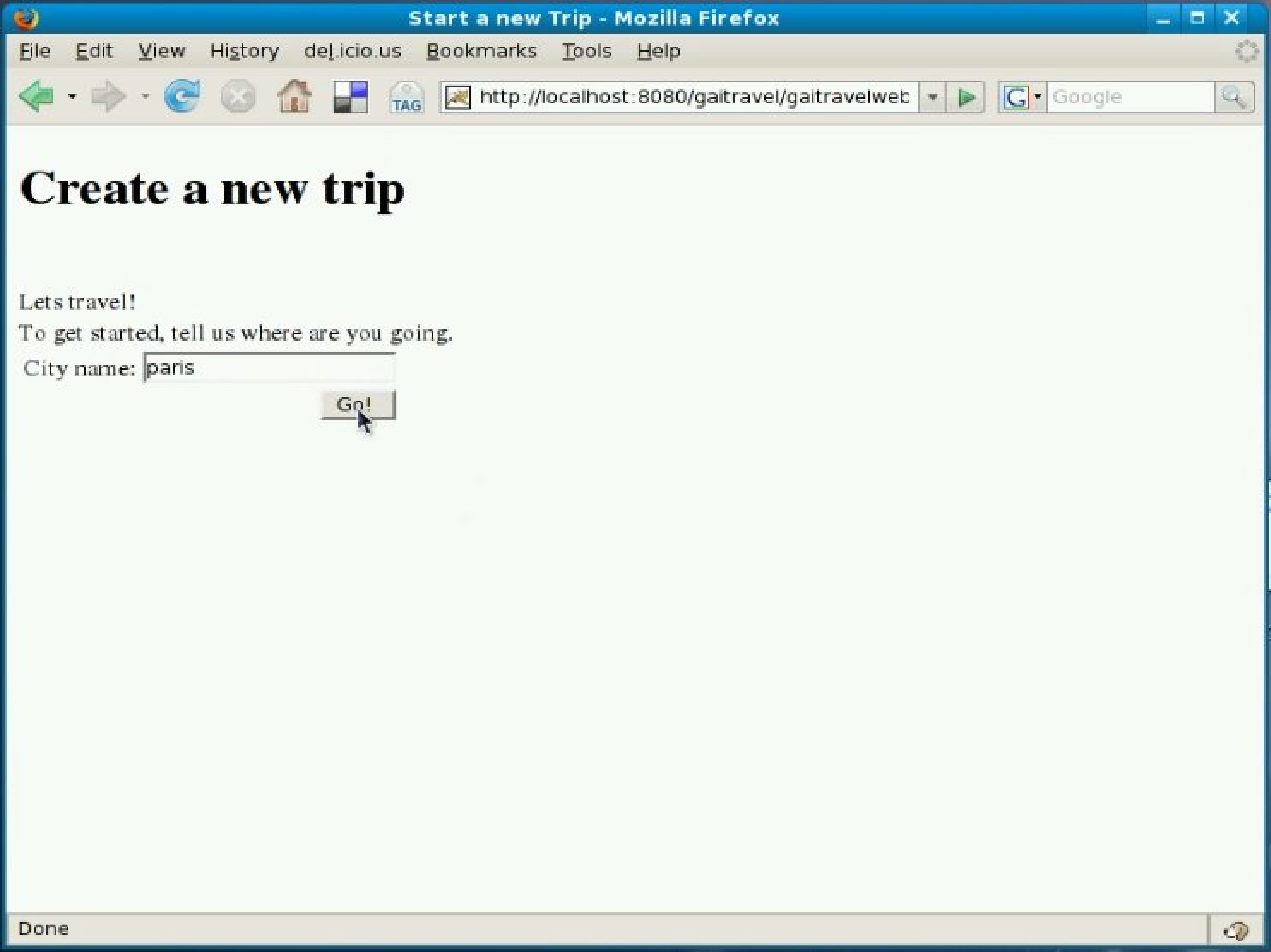
under the constraint $\sum_{j=1}^{n} w_j x_j \leqslant c,$

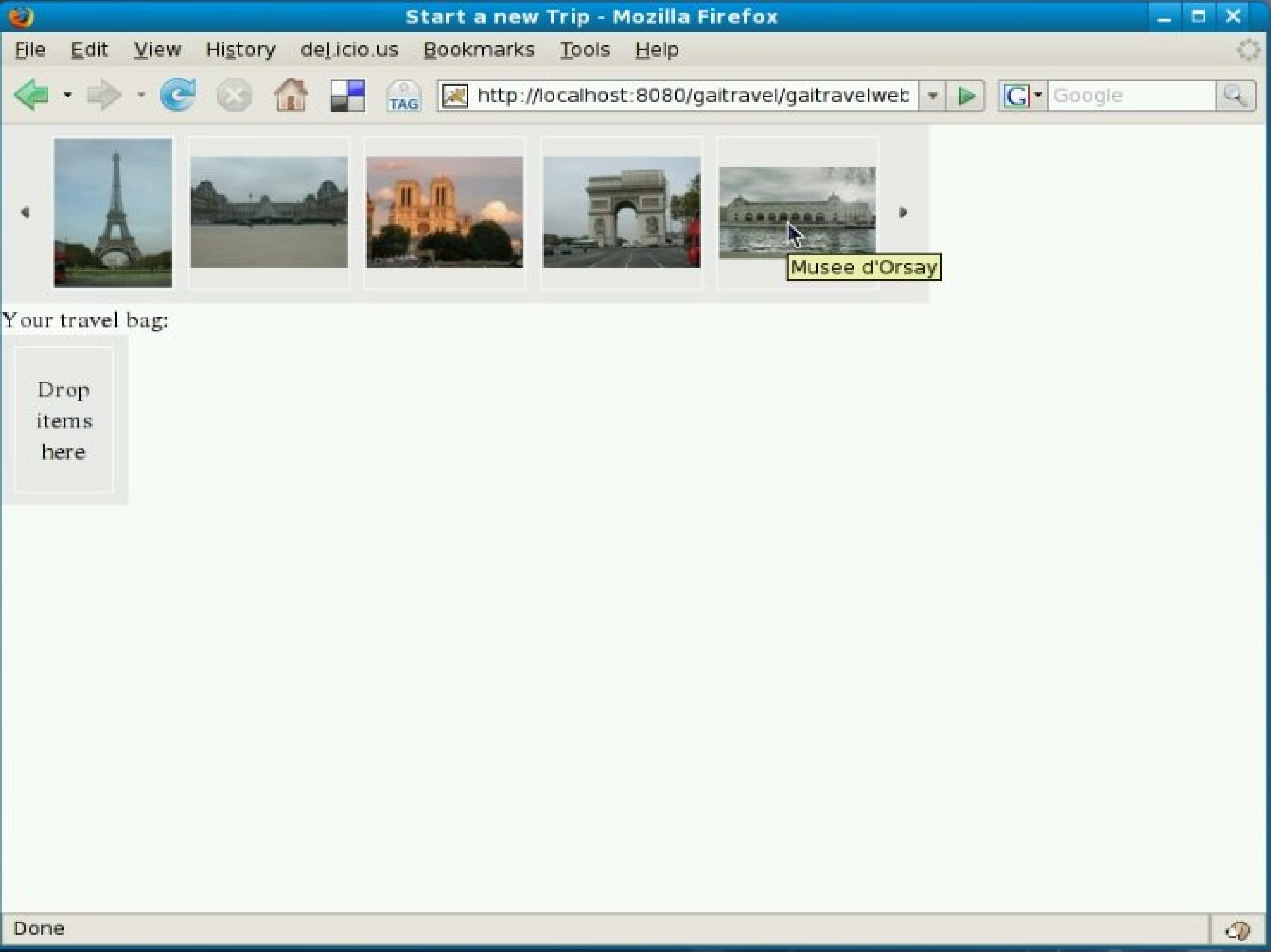
GAI-decomposable function

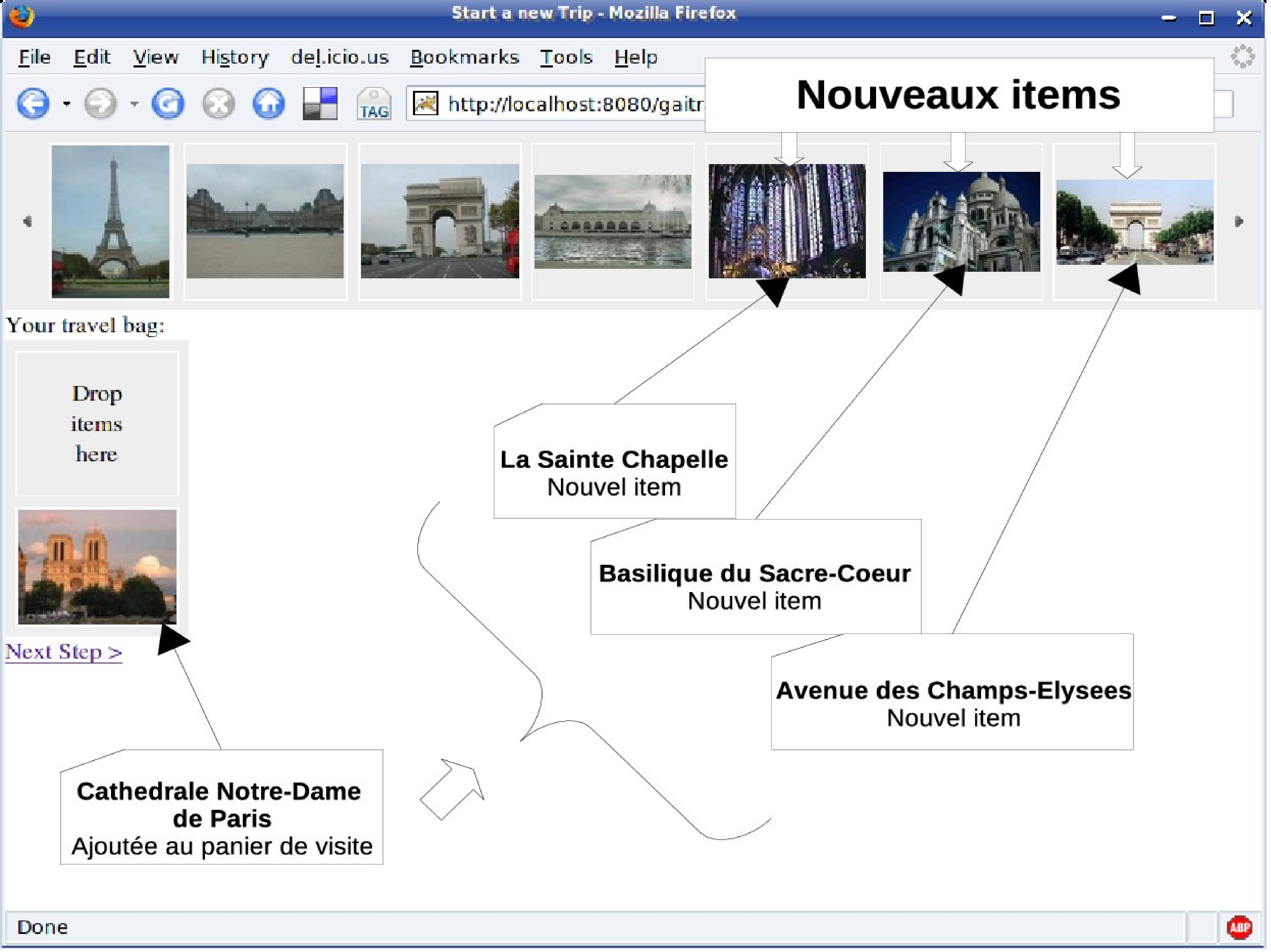
- u_i is the utility function for the user preferences
- w_j is the time needed to visit item j
- x_j in {0, 1} (visit or not item j)
- c total time available in the trip

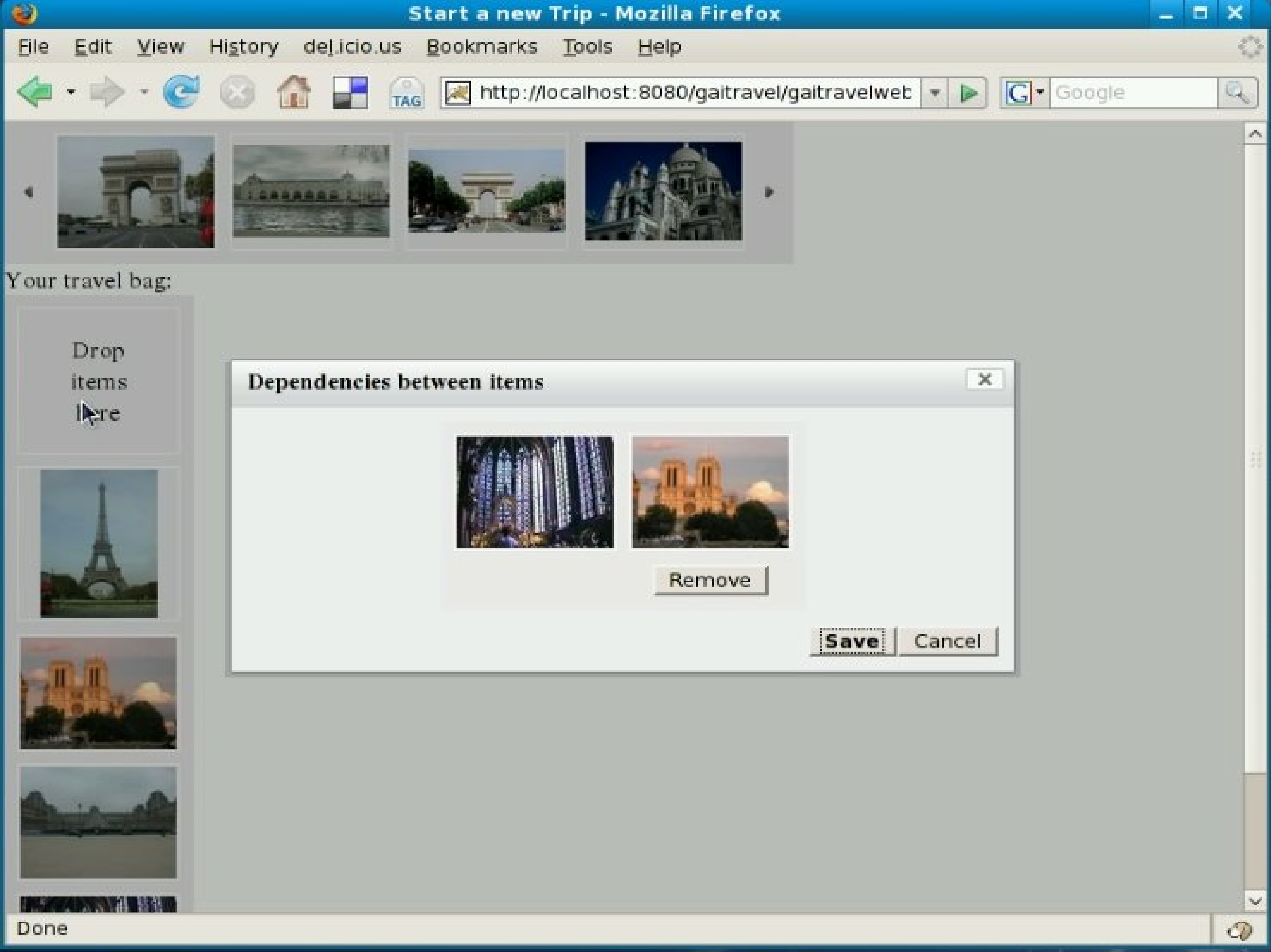
We developed an efficient procedure that uses GAI-networks to solve this kind of knapsack problem

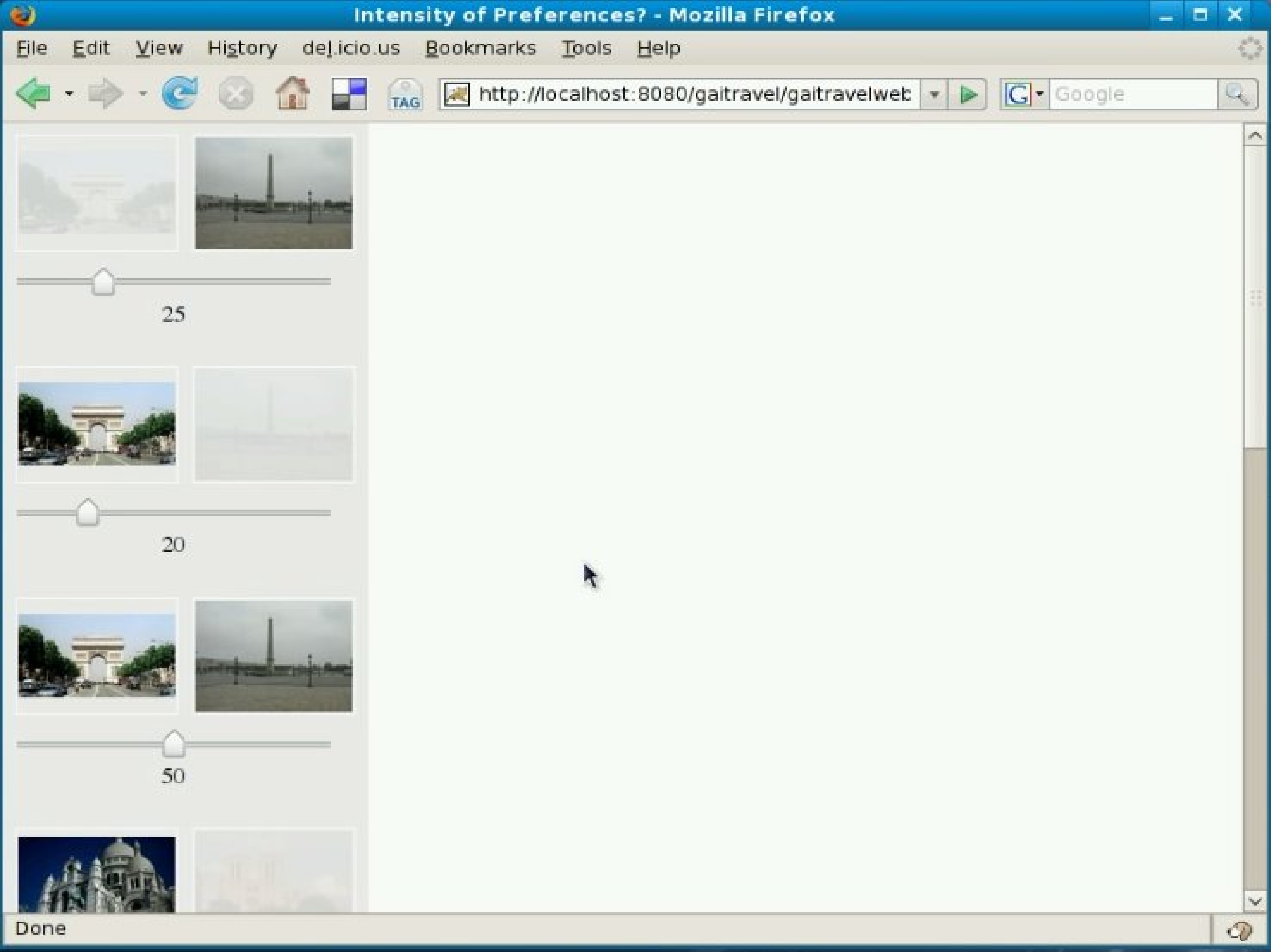
Gvisite? A real application

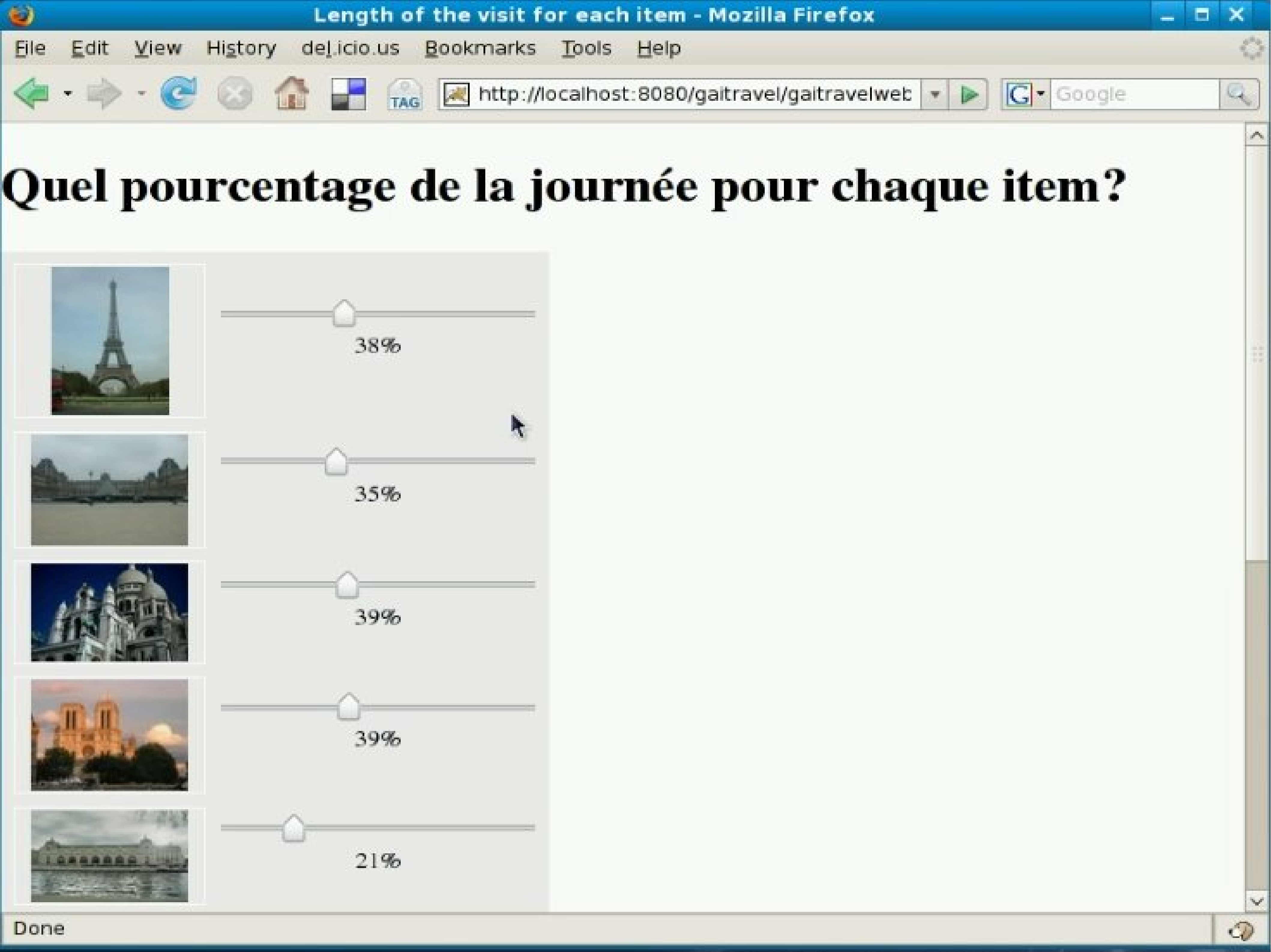

Start a new Trip - Mozilla Firefox
File Edit View History del.icio.us Bookmarks Tools Help
http://localhost:8080/gaitravel/gaitravelweb
Google
Create a new trip
Lets travel!
To get started, tell us where are you going.
City name: paris
Go!
Done

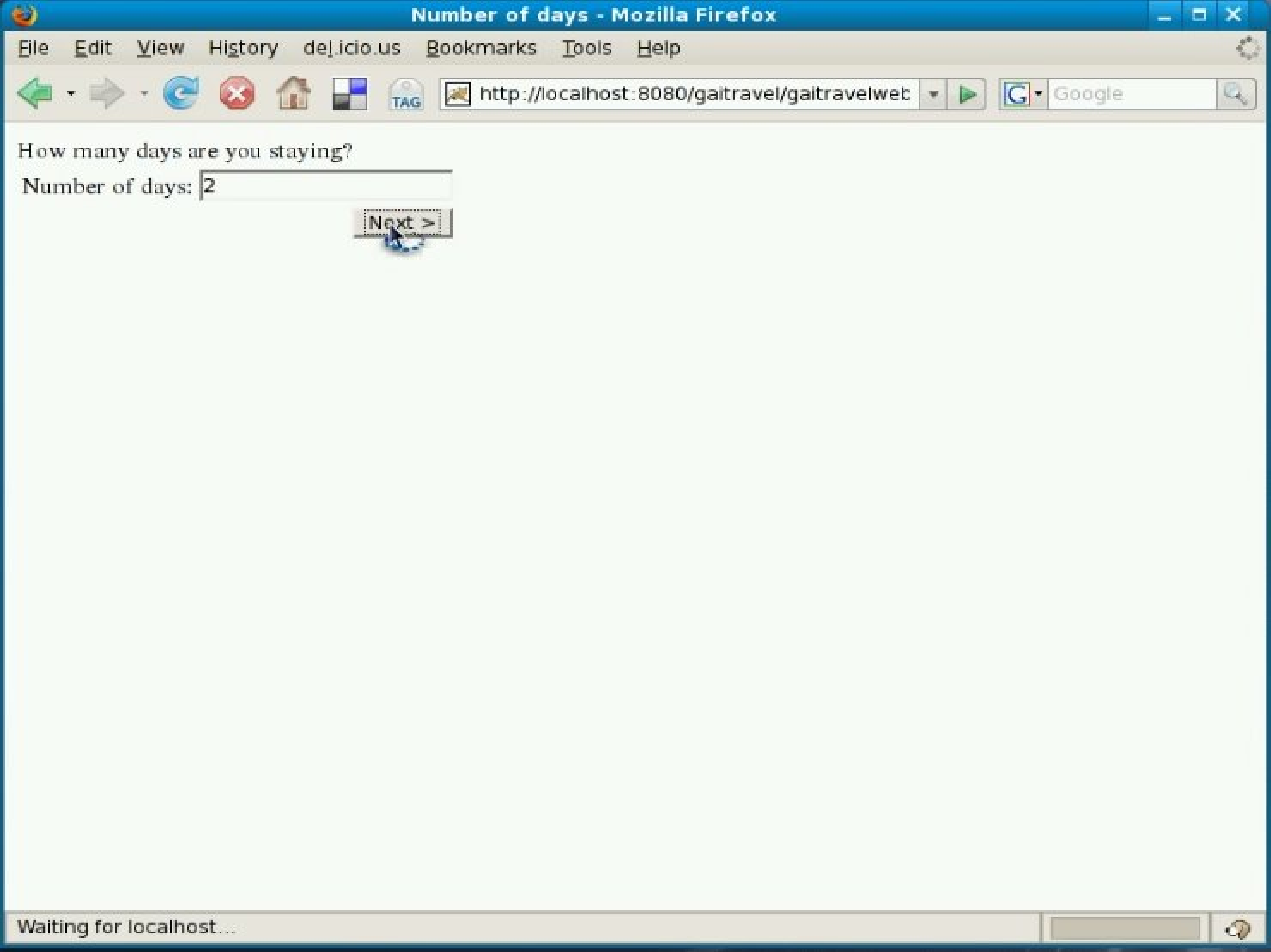

Start a new Trip - Mozilla Firefox
File Edit View History del.icio.us Bookmarks Tools Help
http://localhost:8080/gaitravel/gaitravelweb
Google
Musee d'Orsay
Your travel bag:
Drop items here
Done


Start a new Trip - Mozilla Firefox
File Edit View History del.icio.us Bookmarks Tools Help
http://localhost:8080/gaitr
Nouveaux items
Your travel bag:
Drop items here
Next Step >
La Sainte Chapelle
Nouvel item
Basilique du Sacre-Coeur
Nouvel item
Avenue des Champs-Elysees
Nouvel item
Cathedrale Notre-Dame de Paris
Ajoutée au panier de visite
Done


Start a new Trip - Mozilla Firefox
File Edit View History del.icio.us Bookmarks Tools Help
http://localhost:8080/gaitravel/gaitravelweb
Google
Your travel bag:
Drop items here
Dependencies between items
Remove
Save Cancel
Done


Intensity of Preferences? - Mozilla Firefox
File
Edit
View
History
del.icio.us
Bookmarks
Tools
Help
TAG
http://localhost:8080/gaitravel/gaitravelweb
Google
25
20
50
Done


Length of the visit for each item - Mozilla Firefox
File Edit View History del.icio.us Bookmarks Tools Help
http://localhost:8080/gaitravel/gaitravelweb
Google
Quel pourcentage de la journée pour chaque item?
38%
35%
39%
39%
21%
Done


Number of days - Mozilla Firefox
File Edit View History del.icio.us Bookmarks Tools Help
http://localhost:8080/gaitravel/gaitravelweb
Google
How many days are you staying?
Number of days: 2
Next >
Waiting for localhost...

Solution - Mozilla Firefox

File Edit View History del.icio.us Bookmarks Tools Help

http://localhost:8080/gaitravel/gaitravelwet

For your 2 days in Paris, you should visit:

Catacombes (Les)

Done

Perspectives

- Development of methods that use more natural elicitation questions
- Take care of the evolution of preferences in this kind of model
- Collective recommendation
  - Privacy
  - Non-manipulability