Materializing the Query with Facet-Streams – A Hybrid Surface for Collaborative Search on Tabletops

Hans-Christian Jetter[1], Jens Gerken[1], Michael Zöllner[1], Harald Reiterer[1], Natasa Milic-Frayling[2]
[1]University of Konstanz, Universitätsstraße 10, 78457 Konstanz, Germany
[2]Microsoft Research Ltd, 7 J J Thomson Ave, Cambridge CB3 0FB, UK
{jetter,gerken,zoellner,reiterer}@inf.uni-konstanz.de, natasamf@microsoft.com

ABSTRACT

We introduce "Facet-Streams", a hybrid interactive surface for co-located collaborative product search on a tabletop. Facet-Streams combines techniques of information visualization with tangible and multi-touch interaction to materialize collaborative search on a tabletop. It harnesses the expressive power of facets and Boolean logic without exposing users to complex formal notations. Two user studies reveal how Facet-Streams unifies visual and tangible expressivity with simplicity in interaction, supports different strategies and collaboration styles, and turns product search into a fun and social experience.

Author Keywords

Collaborative Search, Tabletop, Tangible User Interfaces.

ACM Classification Keywords

H5.3. Information interfaces and presentation (e.g., HCI): Group and Organization Interfaces.

General Terms

Design, Human Factors, Experimentation.

INTRODUCTION

The use of tabletops for co-located collaborative search is an ongoing topic in HCI research [18]. Tabletops can offer diverse benefits and potentials for collaborative search such as a closer face-to-face collaboration and more equitable working style [22], an increased awareness and better group work experience [1], and a horizontal form-factor whose affordances are well-suited to follow-up activities (e.g. sorting, sensemaking, making a purchasing decision) [1,18]. However, other potentials of tabletops for search are still unexplored, e.g. the use of "hybrid surfaces" like [14,25] that use tangible interaction with physical props in combination with multi-touch [15]. Except a single design study for video search [11], such hybrid tabletop interaction has not been used in search scenarios yet. Furthermore, in the light of the popularity of tabletops (e.g. Microsoft Surface) in showrooms or flagship stores, it is surprising that no prior research has focused on the obvious task of collaborative search for products in a retail environment. In this paper we therefore present Facet-Streams (Figure 1), a novel design for collaborative faceted product search. It uses a hybrid interactive surface that combines information visualization techniques (a filter/flow metaphor [28]) with tangible and multi-touch interaction to materialize collaborative search on a tabletop. Thereby, unlike in most previous work, our notion of search does not mean to populate an empty workspace with the results from a keyword search. Instead we mean a process of faceted collaborative filtering of a product catalog until the amount of results is sufficiently small to review and decide [10]. Furthermore, in a retail environment like a flagship store a "good" customer experience with "soft" factors such as fun, innovative design and social experience is often valued over "hard" factors such as task completion times and rates.

Figure 1. Facet-Streams on a Microsoft Surface tabletop.

Our work has therefore been guided by three research questions: (Q1) Does our design turn collaborative product search into a fun and social experience with increased group awareness? (Q2) Can we support the great variety of different search strategies and collaboration styles in different teams with a simple but flexible design? (Q3) Can we harness the expressive power of facets and Boolean logic without exposing users to complex formal notations? In the following, we discuss related work and the specifics of our context of use. Then, we introduce Facet-Streams and the underlying design rationale. We describe two user studies and discuss their results in terms of user experience, collaboration styles, and awareness. We conclude by summarizing our results and discussing them with respect to our research questions.

RELATED WORK

Our work roots in the state-of-the-art from three fields of HCI research: (1) tabletop interfaces for collaborative search, (2) tangible queries, and (3) visual query languages.

(1) Morris et al. provide a comprehensive overview and analysis of tabletop search systems along dimensions such as search input, collaboration style and application domain [18]: Regarding search input, Facet-Streams is the first approach that uses hybrid surfaces with tangible and touch interaction. All previous approaches entirely rely on touch, mouse, or keyboard input without making use of any physical props as tangible user interface elements. Regarding the collaboration style, Facet-Streams is similar to FourBySix Search [9], Cambiera [13], and WeSearch [19] which all support seamless transitions between tightly-coupled collaboration and loosely-coupled parallel work. However, unlike these applications, Facet-Streams does not use keyword search for Web, document, or multimedia retrieval but uses a visual and tangible query language for faceted search. Thus, Facet-Streams shares commonalities with TeamSearch that also creates a faceted search experience based on Boolean-style AND queries on tagged photo collections [17]. Like TeamSearch, we use circular widgets to specify categorical criteria (or facets) but aim at a far greater query expressivity with arbitrary numbers and logical combinations of such widgets including AND and OR. Furthermore, we do not restrict users to only formulate either personal queries or collective queries. Instead we want to enable them to develop multiple personal and collective queries in parallel and to freely shift criteria between them for maximum flexibility in strategies and collaboration styles. A further fundamental difference between previous work and our design is the employed notion of search. Except TeamSearch and PDH [24] all systems in [18] increasingly populate the collaborative workspace with the results of keyword searches. Thereby search has the notion of adding result sets to the shared workspace. In contrast, we want to follow a faceted search approach [20] where search means narrowing down the entirety of products in the workspace to the desired subset. Thus the focus of collaboration in Facet-Streams lies on the formulation and logical combination of the desired facets of a product (e.g. "price < 100") *before* reviewing individual results. This is opposed to related work where collaboration is focused on reviewing and relating results *after* search. The only systems in [18] following a similar faceted approach have either limited expressivity (TeamSearch) or force users to only navigate a single facet at a time (PDH).

(2) Ullmer et al. were the first to suggest physically constrained tokens to manipulate database queries and result visualizations [27]. Two kinds of physical tokens (knobs and range sliders) served as tangible input controllers that are put into slots next to a display. Although enabling some basic Boolean logic between the range sliders, the overall expressivity was limited to assigning database fields to the axes of a scatter plot and altering parameters of predefined queries. Nonetheless, this design inspired many further designs of tangible queries (e.g. for facilitating search for children [6]) or Blackwell et al.'s Query by Argument (QBA) system [3]. QBA enables groups to manifest the course of an argument in spatial configurations of "statement tokens", i.e., RFID-tagged cards as place-holders for contributions to the discussion. Each token carries a reference to a virtual information item that contains the contribution (e.g. relevant text passages). By spatially configuring the statement tokens during discussion, the group provides continuous relevance feedback to an information retrieval system that continuously evaluates the spatial structure to adjust its ranking mechanisms. As a result, the system suggests related material on a peripheral screen. QBA's approach to use spatial configurations of tokens to materialize the (chrono-)logical order of an argument during a collaborative process has been inspirational for our use of a network of tokens for faceted search. However, we want to provide users with tokens for precise filtering and immediate feedback. QBA is targeted at working invisibly in the background to gradually adjust its ranking without the same need for precision and immediacy.

(3) While being of great practical value for search, Boolean AND and OR are concepts difficult to grasp and they contradict the linguistic sense of "and" and "or" in our natural language use [4]. Therefore many attempts have been made to visualize these concepts in visual query languages and interfaces: Today's faceted search on e-commerce Web sites [20] and faceted visualizations such as [5,16] allow to formulate the equivalent of sophisticated Boolean queries by taking a series of small, simple navigation steps. However, these designs are for single users only and do not provide a random access to all intermediate steps in navigation history. This hampers their use for co-located collaborative search where all criteria of multiple parallel queries must be accessible for iterative refinement at all times. Young et al.'s filter/flow metaphor [28] achieves this by visualizing a Boolean query as a sequence of logically linked nodes that carry the criteria. However, it only permits a single query per workspace and its simple layout uses too much screen estate for tabletops. FindFlow [8] expands this metaphor into a 2D plane with a more efficient use of screen estate, but lacks parallel queries and the Boolean OR for specifying complex criteria (e.g. "either the hotel has a good restaurant OR I want to have a small kitchen in my room."). It also conflicts with the size limitations and legibility around a tabletop. This is also true for DataMeadow that uses a filter/flow metaphor for connecting data, filters, and visualizations for visual analytics [7]. Similarly LARK uses a filter/flow metaphor for managing multiple-coordinated views for collaborative visual analytics on large multi-touch displays or tabletops [26]. However, in contrast to Facet-Streams, both systems are not used for Boolean search and they also do not employ any tangible user interface elements.

Figure 2. Semantic Zooming in Facet-Streams: grid view of hotels (left), zooming into details (center), index card (right).

THE DESIGN OF FACET-STREAMS

Our design is based on a scenario of use in which three family members gather around a Microsoft Surface in the showroom of a travel agent. They want to find and agree on one hotel for a week of vacation in Europe. Each of the 204 hotels in the catalog carries 12 quantitative or categorical facets (e.g. *price per night, recommendation rate, hotel features, country*). For every hotel each of these 12 facets has a single value (e.g. Hotel A has *price per night* = 50 EUR; *recommendation rate* = 80%; *hotel features* = WiFi+Gym, etc.). We have used this scenario throughout our entire design process and for our user studies. To ensure that we work with realistic product data, the online travel agency holidaycheck.com provided us with a subset of their catalog including authentic user photos and ratings.
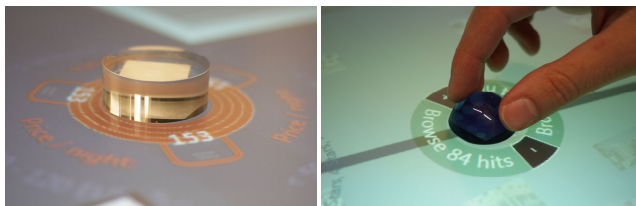

Figure 3. A *facet token* (left) and a *result token* (right).

The starting point of our design has been a simple grid layout on a zoomable canvas that contains all hotels (Figure 2). We refer to this initial state of the system as *browsing mode*. In this mode, users can access the details of a hotel by semantic zooming into its icon using typical multi-touch manipulations (tapping or pinching for zooming, sliding for panning). Users are enabled to collaboratively browse a catalog of products at different levels of detail while maintaining spatial orientation and using visual cues or thumbnails for recognition. However, this *browsing mode* is hardly efficient for a search scenario in which various criteria have to be met: While family member A is primarily interested in a close beach, family member B prefers a place with good service and food, and family member C is concerned about the travel budget. Finding a hotel that suits all these criteria by manually reviewing each hotel would be a too tedious and error-prone task.

Using Physical Tokens for Faceted Search

To leverage the power of collaborative faceted search, we enable each user to specify a set of personal criteria (e.g. *country* = Spain and *price* < 120 EUR) for narrowing down the number of displayed hotels. This is done using physical facet tokens (Figure 3). Each token is a circular glass disc of 30 mm x 12 mm. Each token carries a unique rectangular fiduciary marker (19 mm x 19 mm) on its bottom side. The id, X and Y positions and orientations of all markers are tracked by the tabletop's vision system. The choice of the form and material of the token is not random, but is intended to create affordances and make appropriate actions perceptible to the user [21]. The token's shape resembles that of a piece on a checkers board or an ice hockey puck and thereby affords sliding the token with the flat side lying on the tabletop. Its diameter invites to comfortably grab the token between thumb and index finger and enables rotational movements of sufficient precision. Thus, the token's physical form already hints at its manipulation possibilities of translation and rotation. This also helps to prevent unanticipated uses such as leaning tokens or rolling tokens over the table in an upright position. The choice of glass is based on previous observations of issues of hygiene in tangible and touch interaction [23]. We hoped that glass tokens would be considered as more hygienic than plastic or wood by those who feel uncomfortable with touching a shared object. Furthermore, future designs could explore the use of transparent markers printed with infrared-absorbing ink, so that the glass token itself could be used as a small screen for displaying further information.

After a *facet token* is put on the tabletop, the system switches into *query mode*. A new translucent layer appears above the canvas, the canvas freezes, its colors are dimmed, and the screen zooms out to display the entirety of hotels in the background of the new layer. On the new layer, virtual elements are displayed around each *facet token* to augment them with content and functionality (Figure 4). Users use a *facet token* to formulate a criterion by selecting a *facet* (e.g. *hotel stars*) and specifying the desired values (e.g. *hotel stars* = 1 or 5). For changing the facet, users touch the currently selected facet that is displayed along the inner circle around the token. This opens a circular pop-up menu called *facet wheel* in which a single facet can be selected by touching its wedge. For changing the desired value range, users touch the currently specified range that is displayed along the outer circle around the token. This opens the *value wheel* in which values can be specified by selecting or deselecting value wedges by touch. The value wheel either carries wedges with values for categorical facets (e.g. CableTV, Restaurant) or quantitative value ranges (e.g. < 50 EUR, 50 – 80 EUR).

Figure 4. A facet token with *hotel stars* = 1 *or* 5 as criterion (left). Touching the current facet ("Hotel Stars") will invoke the *facet wheel* (center). Touching the current value ("1 Stars, 5 Stars") will invoke the *value wheel* (right).

Collaboration Styles: Collective vs. Personal Criteria

Using our *facet tokens*, we could have created a faceted search similar to TeamSearch [17] in a straight-forward way: By combining the criteria from all *facet tokens* on the tabletop with a logical AND, we would have provided a simple mechanism to formulate a collective query. However, as mentioned before, we wanted to introduce a greater degree of flexibility in search strategies and collaboration styles, which is of great importance for collaborative search [18] and in particular in our case of product search. A collective query based on a logical AND of all criteria affords very tightly-coupled collaborations. Each modification of a single criterion has an immediate effect on the entire result set. This can be disruptive in earlier phases, when users will start with unrealistic and conflicting criteria (e.g. a very luxurious hotel for a very low price) because they did not have the chance to individually explore the available facets and products first. The search process can then quickly deteriorate to a random trial-and-error manipulation of criteria to receive non-empty result sets. Therefore, we wanted to enable phases of parallel personal exploration. Even in later stages, such phases can serve users to verify and alter their personal contribution to the collective criteria. Therefore seamless switching between tightly-coupled collaboration and loosely-coupled parallel work is crucial.

Networks of Facet Tokens as Query Language

As already discussed, we fundamentally redesigned the visual filter/flow metaphors for Boolean logic of [8,28] to enable collaborative faceted search. The first step of our redesign was to permit multiple unconnected chains or networks for parallel collaborative exploration. The second step of redesign permitted their use on tabletops by improving the use of screen estate and largely reducing the visual complexity and the amount of text output. A fundamental concept in our redesign are *streams*. A s*tream* visually and logically connects two physical *facet tokens* (Figure 5). Once connected by a stream, tokens will not lose this connection, even when moved to different or distant locations on the tabletop. The stream connects two tokens until one of them is physically removed from the tabletop or the user manually cuts the stream with a touch gesture. While new to the search domain, similar designs are used on tabletops for live performances of music or video

synthesis [14,25]. In our design, streams are directed connections of criteria for faceted search: the source token emits an output stream of filtered hotels that is received as input by another token. The output stream of a token only contains those hotels from its input stream which meet the token's criterion. Thus, if the token's criterion is *country* = Spain, only those hotels from the input stream that are located in Spain will be emitted into the output stream. If a token has no input stream, it is treated as one of many possible sources or starting points in a network and applies its criterion on the entirety of all hotels in the catalog.

Multiple tokens can be connected to chains of criteria where each token only forwards the hotels from the input to the output that match the token's criterion. Therefore these chains of tokens are an equivalent of a Boolean expression where all criteria are combined using a logical AND. To further extend the expressivity of our query language, tokens can have multiple output and input streams. The output streams are all identical and allow users to stream the output towards multiple tokens in parallel. The multiple input streams are combined internally using a logical AND (Figure 5). However, throughout design and user testing, we strictly avoided to use any mathematical or linguistic formalism to convey this Boolean logic on the interface or during instructing users. Our design goal has been to move these concepts onto an entirely visual layer of reasoning.
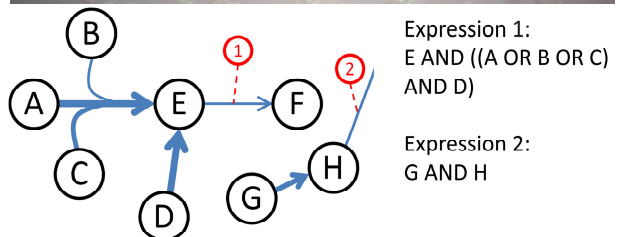

Figure 5. A network of *streams* that connect *facet tokens* (top). The Boolean equivalent of the networks at ① & ② (bottom).

We have also integrated the possibility to let two streams flow together: By directing an output stream onto an existing target stream, the output stream is bent from a straight line into a Bezier curve that flows into the target stream (Figure 5). Internally the flowing together of two streams is treated as a union of both streams based on a logical OR. The possibility to let streams flow together increases the expressivity of Facet-Streams. The example of a user who wants either a great restaurant or an own kitchen in the room can be covered by letting two according output streams flow together. The resulting union can again be used as input for a further chain or network of criteria.

To better support users' understanding of streams, we have integrated immediate visual feedback that is updated instantly after any user input: The number of results flowing through a stream is logarithmically mapped to the thickness of a stream. Empty streams are shown as thin lines, but are highlighted in red color to differentiate them from streams with few results. As additional feedback the number of outgoing results from a token is numerically displayed around it. Furthermore, when selecting new value ranges inside a value wheel, each value wedge indicates how many results will remain in the output stream after (de)selection. To get a quick overview of the results flowing in a stream, users can also touch the stream at any time for an immediate query preview. All hotels that are not contained in the stream then disappear from the zoomable canvas in the background and a numerical value with the number of the remaining results appears close to the finger. This kind of feedback can also be achieved by putting a *result token* on a stream (Figure 3). The token additionally provides easy ways to browse sequentially through all contained results or to temporarily switch back into the *browsing mode* for freely exploring the remaining result set.

Designing for Low Viscosity and Parallel Interaction
To enable users to collaboratively explore multiple queries and to shift criteria between them, we have designed the lower level interaction techniques following our design principles of *low viscosity* and *parallel interaction*.

Our understanding of *low viscosity* is based on Blackwell et al.'s cognitive dimensions of notations framework [2] and for us means a "low resistance to change". This is crucial, since our support for different search strategies and collaboration styles depends on the ability of the users to quickly adapt the topology, the spatial layout, and the criteria of the networks according to the intended working style and goals. For achieving *low viscosity*, the use of a hybrid surface is of great benefit: By coupling the virtual representation of the network to physical props, rearranging the spatial layout is an entirely physical activity without the need for using pointing devices or learning touch gestures. The physical tokens can be relocated by carefully dragging them, but also by carelessly sliding them or even wiping them off the tabletop with the arm. For quick explorations of alternative network topologies, tokens can be lifted from the table and

put at arbitrary locations or also into streams. Thus, the manipulation of the tokens' locations entirely takes place in the physical not in the virtual world and does not create an uneasy feeling or fear of irreversibly damaging or destroying virtual content. However, to achieve this, many details of the interaction had to be considered. For example, we store the selected facet and value range of a *facet token* even after it is removed from the tabletop, so that temporarily lifting or relocating a token does not destroy the contained user settings. Similarly, when removing a token from a network, a set of rules is applied that ensures that all neighboring tokens which have been disconnected by the removal are reconnected in a sensible way. The surrounding network topology is not destroyed. For other frequent manipulations, we decided to use direct touch instead of tangible interaction: Streams can be created by touching one of the orange handles that are displayed around the tokens (Figure 4). They can then be directed towards their destination with sliding the finger. For cutting streams, a crossing gesture with the finger is used. This direct touch interaction was employed to increase precision and efficiency and also to create a consistent separation of concerns: The position and orientation of the physical tokens are changed using tangible interaction. The creation and cutting of the virtual streams between them happens by direct touch on the virtual elements augmenting the tokens.

The design for *parallel interaction* supports different collaboration styles by enabling users to better parallelize work. For example, a single user might want to modify criteria in a personal network while two other users are working together on a different shared network. Therefore the interface must be able to handle simultaneous tangible and touch input from all involved widgets such as *facet* or *value wheels* without concurrency issues. Modal interaction or global interface modes must be avoided. Furthermore, all widgets or elements must be movable to quickly establish temporary personal or shared regions on the tabletop, so that parallel user tasks do not interfere with each other. In our design, all widgets and elements are therefore attached to physical tokens to effortlessly control their position. To ensure accessibility and legibility from all sides, we use a circular design for all widgets to achieve a more equitable interaction without a preferred orientation. Wherever possible, interactive elements or labels around the tokens appear three times every 120 degrees and are curved around the edge of the circular token (Figure 4). A further design for *parallel interaction* is to couple the orientation of a widget to that of its physical token. Thus, rotating the token allows effortless changes of orientation without the need to touch virtual handles or to apply touch gestures for rotation. This enables users to quickly show or pass a widget to a collaborator with a different viewing angle. This also enables bimanual interaction of a single user: One hand can be used to rotate the token of a value wheel while the other hand is used for touching the wheel without lifting the finger. This selects the entire value range that passes below.
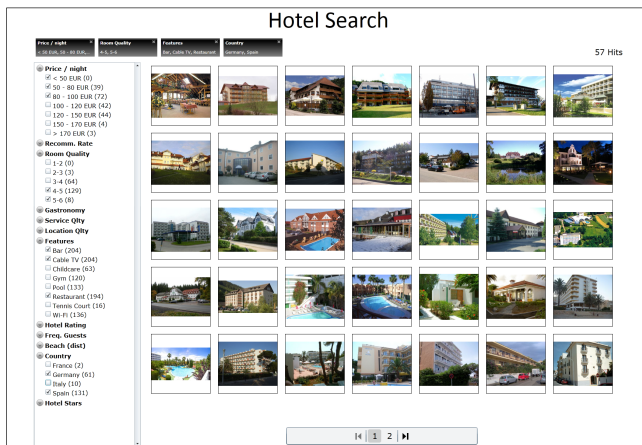
Figure 6. The Web interface for faceted hotel search.

USER STUDIES

We conducted two user studies with different foci to get insight about the usability and utility of Facet-Streams and to collect empirical data for answering the research questions that we have formulated in the introduction.

Study 1: Comparative Study of Collaborative Use

In the first study we wanted to observe the use of Facet-Streams during a realistic collaborative search task in which a compromise had to be negotiated by the group members. We were interested in how participants would interact with the interface in terms of search strategies, collaboration styles, and parallel interactions. A Web interface for faceted search served as control condition (Figure 6). We compared the interfaces in terms of the objective quality of group compromises and collected qualitative data to contrast the observed personal and group behaviors and strategies.

Design & Participants. We used a between-subjects design (IV: interface type Facet-Streams or Web interface) with 72 participants, randomly assigned to 24 groups of three. We opted for a between-subjects design as we identified two aspects that can have a significant and uncontrollable influence in a within-subjects design: First, the novelty of a tabletop with tangibles might evoke a strong "wow"-effect and great bias when put in contrast to a traditional desktop interface. Second, group dynamics evolve over time as people get to know each other. Even a counter-balanced within-subjects design might not be able to rule out interaction effects. Our participants were students or faculty from a variety of non-technical subjects (only two students of computer science). The average age was 25 (SD = 7.4 years) with 36 females and 36 males.

Tasks & Procedure. Our tasks required participants to agree on a single hotel from a set of 204 hotels within a limited amount of time. In order to simulate a realistic scenario, participants had to agree on compromises and make concessions to solve a task: For each task every group member was assigned three personal criteria (e.g. 'the hotel has to have 3 or 4 stars', 'the distance to the beach must be smaller than 150 m', 'the room quality must be 4 or 5').

However, the task was designed by us in a way that made it impossible for the group to satisfy the total of nine criteria simultaneously. Simply combining all criteria with a logical AND always led to an empty result set. Thus, all participants had to negotiate whose personal criteria to soften (e.g. by extending the price range) or which to give up completely. The whole group was instructed to find an "optimal" compromise that is as close to the entirety of all 9 personal criteria as possible. We gave participants three of these tasks with varying difficulty in terms of conflicting criteria. A soft time-limit of 7 minutes per task was used to limit and control the session duration and increase participants' motivation to come to a decision. However, we did not interrupt users before a final decision was made, since the time limit was not intended as a sharp criterion for task completion or failure. The mean duration of a task has been 6:41 min (mean = 401s, SD = 133s). Prior to working on the task, each group was given a five minute instruction to the system and five minutes for free exploration. After completing the three tasks, each participant filled out a personal questionnaire about their subjective assessment of the system. Each session took about 45 minutes and was video recorded from different angles to grasp not only the interaction with the interface but also the group dynamics. Participants were compensated with 15 EUR for their time.

Interfaces. We used Microsoft Silverlight to design a Web interface that matches the state-of-the-art of faceted search on large e-commerce sites (e.g. Amazon.com). Our design was slightly more advanced than traditional interfaces since animations and dynamic queries were used to create a more responsive rich internet application. To replicate today's reality of collaborative search, participants were asked to solve tasks on one 24" screen sharing one keyboard and mouse. To enable a fair comparison, we used a pre-test to identify specific features from both interfaces that would give users the ability to "cheat" our test design. As a result, we removed the continuous update of the remaining hotels from the checkboxes of the Web interface and from the wedges in the value wheel of Facet-Streams. Furthermore, we removed the logical OR from Facet-Streams, as the Web interface does not have an equivalent functionality.

Study 1: Results

We first analyzed the objective quality of the results that the participants achieved with both systems. We compared the results for each task from 12 Web interface groups and 11 Facet-Streams groups with the optimal result for each task from the catalog. We had to exclude one Facet-Streams group from this analysis, as one participant repeatedly ignored the task instructions and used only his personal real-life preferences to judge the group's results. For analysis, we determined the distance between the given criteria and the selected hotel for each group. We distinguished between concessions and fails. A concession means that the criterion is not met by the hotel, but is met by one of the neighboring values. Each concession adds 1 to the distance. A fail means that the criterion is neither met

by the hotel nor by neighboring values (e.g. the given price range was 50-80 EUR and the hotel costs more than 120 EUR). A fail adds 3 to the distance. Deviations in categorical facets such as country were treated as fails. Deviations in features were counted as one concession per missing feature. Table 1 shows the mean distance for each task. Differences between Facet-Streams and the Web Interface were statistically not significant (T1: $t(21)=-0.847$, $p=0.407$; T2: $t(21)=0.638$, $p=0.531$; T3: $t(21)=-1.517$, $p=0.144$). As all tasks made concessions necessary, the "optimal" value is given as well.

	Facet-Streams	Web Interface	Optimal
T1	2.00 (SD = 1.00)	2.33 (SD = 0.89)	1
T2	6.64 (SD = 1.29)	6.33 (SD = 0.98)	4
T3	3.27 (SD = 0.90)	4.16 (SD = 1.74)	2

Table 1. Mean distance and optimal distance for T1-T3.

In comparison to the optimal result, groups achieved good compromises on average with only 1 to 3 concessions or 1 fail per task. Thus, both interfaces proved to be effective for the given collaborative search task. This is also confirmed by the subjective assessments of the participants from the questionnaires. On a scale from 1 ('is not true') to 7 ('is true') the mean scores per group for "The system has supported us well for the tasks." are 6.13 (SD = 0.43) for Facet-Streams and 5.86 (SD = 0.77) for the Web interface. While the mean scores are slightly in favor for Facet-Streams this is not statistically significant. Significant differences in favor of Facet-Streams exist concerning the fun users had while using the system and the perceived innovativeness of the design. The mean score for "I had fun working with the system." for Facet-Streams is significantly higher than for the Web interface ($6.69 > 5.69$, $t(23) = 4.716$; $p < 0.001$). The same is true for "The system is very innovative." ($6.38 > 3.61$, $t(23) = 8.444$; $p < 0.001$).

Usability and User Experience. We find it notable that the users of both systems achieved an equal objective quality of results. Since Facet-Streams introduces a novel and unfamiliar style of hybrid tangible and touch interaction with a filter/flow metaphor, we had expected that users would have more difficulties to achieve equally good results within the given timeframe. Although there was only a brief period of introduction and free exploration, users quickly mastered Facet-Streams' interface and did not achieve inferior results compared to the groups using established faceted navigation with a mouse. Furthermore, as mentioned above, the Facet-Streams interface was reduced in functionality for a fair comparison (no indication of the number of hotels on wedges in the value wheel, no logical OR). In particular the indicators on the wedges would have most likely improved the quality of results in real usage. With respect to the subjective assessments, participants perceived the design of Facet-Streams as something innovative and fun to work with. This is notable,

as these assessments were made as part of a between-subjects study to reduce bias resulting from the novelty of a design. This notion of a fun experience was also observed during the sessions: After receiving a new task from the experimenter, one participant turned excitedly to his collaborator and stated "Nice game!". A participant in another group repeatedly stated "This is so much fun" and frequently requested more time for "playing around". She was disappointed at the end of the session after realizing that the last task had been solved ("What a pity!").

Search Strategies and Collaboration Styles. Transitions between tightly-coupled collaboration to loosely-coupled parallel work and vice versa were a reoccurring theme in the 11 Facet-Streams groups. From the 33 tasks that were performed by these groups, 26 tasks were begun with a phase of loosely-coupled parallel work. Participants started by building small personal networks of up to 3 tokens in parallel for an initial exploration of their personal criteria. Only after two or more participants had completed this exploration, a phase of more tightly-coupled collaboration with joint networks took place. This hints at the importance of having separate but joinable workspaces. In the 7 other cases, participants started with a tightly-coupled collaboration from the beginning. Here, all participants sequentially added criteria to a collective chain of tokens. However, in two cases, single participants later seemed to feel that this initial strategy is not meeting their needs. They then started to explore their personal criteria with own tokens in parallel without explicitly discussing this change of strategy with their collaborators. Therefore the number of tasks that involved loosely-coupled parallel work totals to 28. During these 28 cases different kinds of transitions from parallel to tightly-coupled work could be witnessed. In 11 cases, all personal networks of the collaborators were merged more or less simultaneously to a single collective query. This was done to identify the hotels that meet all the criteria of the participants and to review candidates that could solve the task. In 12 cases the transition to collaboration happened between pairs of participants who spontaneously decided to review the intermediate results they share. In 8 cases such pairwise merging of networks was employed systematically to compare results and to identify conflicting criteria: Instead of merging all networks to a collective query, only two networks were merged at a time. Only after the conflicts between each network had been solved by pairwise merging, all networks were merged to a collective query solving the task. Thereby it is noteworthy, that merging networks did not interfere with the participants' awareness of which token had been contributed by whom. Whenever a collective network was dissolved into multiple networks again, users easily recognized their tokens because of their spatial position, the network's topology, and the contained criteria. Thus, moving between personal and collective queries was not a challenging task.

Compared to Facet-Streams, the search strategies of the Web interface users appeared much less systematic. Of the 36 tasks that were performed, 21 did not follow any recognizable strategy. In these cases, participants shared their personal criteria verbally with the person who was operating the mouse and who collected and entered all criteria in a random order. In 11 cases, the operator used the list of available facets on the screen to sequentially inquire at every facet whether group members had a relevant criterion or not. In the remaining 4 cases, a similar sequential inquiry took place using the criteria cards that had been handed to the participants at the beginning of the task. Typically the groups encountered a small or empty result set after few entered criteria. From then on the groups started to modify or soften their collective criteria during tightly-coupled collaboration. Not surprisingly, there was intense verbal communication between the operator and the other two participants during all phases, but in particular during this refinement phase. However, in contrast to the Facet-Streams groups, most of this communication happened to check and confirm which criteria had already been entered and which not. Although the region of the Web interface with the checked or unchecked criteria was visible to all collaborators, the groups quickly lost track of the number of entered criteria. Furthermore the participants could not attribute the different checkboxes to individual persons and seemed to be less aware of their own and others' criteria. This led to a lot of "noise" in the verbal communication that primarily served to create awareness for the current system state, but seldom for suggesting actual steps for problem solving. Not surprisingly, this also led to a more browsing-oriented search strategy. Compared to Facet-Streams, manual reviewing of results happened earlier and lasted longer. In conclusion, the interaction with the Web interface suffered from a lack of awareness and from the alienation of the participants from their criteria after being entered into the system. This is contrary to our observations for Facet-Streams and hints at the role that the tabletop plays as a space for creating awareness.

Discussion. Hornecker et al. introduce a framework on physical space and social interaction for tangible user interfaces. They emphasize the importance of *externalizations* (e.g. a shared visual or physical workspace) to support communication, negotiation, and shared understanding. Externalizations can aid cognition, provide shared reference, and remember our traces. They can directly or indirectly foster collaboration and awareness [12]. In Facet-Streams, the shared workspace on the tabletop with networks of tokens served as such an externalization. The networks provided a visual and tangible map of the search process that captured its (chrono-)logical development. Users were able to attribute tokens and their criteria to other collaborators based on the topology and spatial distribution of the networks. They were able to interpret the networks as an indicator of their progress in search and used them to store, revisit, and reuse

intermediate results. These features of the tabletop led to an increased *awareness* that became obvious at several occasions: Frequently, participants reminded each other of their respective criteria. For example, one participant reminded her collaborator to think of the collaborator's children, as the collaborator was about to abandon the requested hotel feature "childcare". Increased awareness also became visible during phases of loosely-coupled parallel work: Although being busy with his own personal network, one participant realized that another collaborator had problems handling the value wheel and quickly interrupted his own work to help out. In some cases, such mutual support involved verbal communication and pointing towards the elements to touch or turn. In other cases, it happened without any prior verbal communication or coordination. For example a participant noticed that one of his collaborators had repeatedly tapped a hotel to review its content. However, the system was currently in *query mode* that does not support this feature. Without any further communication this participant grabbed the close-by result token to switch to *browsing mode*, so that his collaborator's flow of work was not interrupted. Such interactions can be regarded as first steps towards creating a "flow experience" within a group. In contrast, the Web interface's externalization without spatial or topological cues failed to provide this degree of awareness. Users were also less able or more reluctant to provide mutual help. We observed one case in which the operator of the interface made frequent mistakes while (un)checking criteria. However, the two collaborators who closely watched his interactions did not realize these mistakes or decided not to intervene.

Observed Problems. While the *query mode* was well-received, we observed usability flaws during *browsing mode*. For browsing, most groups gathered at a single side of the tabletop, since the orientation of the results could not be controlled individually. In future, users must be enabled to read, arrange and pile results from all sides for an equitable collaborative browsing. Furthermore, users often had to move tokens out of the tabletop's center, since they occluded results or felt too distracting. Finally, the division in *query* and *browsing mode* led to confusion because the current mode and its available functions were not indicated clearly. In future designs we will try to minimize occlusion and to avoid different modes by merging their functionality.

Study 2: Comprehensibility of Filter/Flow Metaphor
We conducted a second study to evaluate the comprehensibility of our metaphor for Boolean logic. We invited 7 subjects to single user sessions (4 male, 3 female, avg. age 23.7 years, SD = 2.75). Participants with a background in computer science or mathematics were excluded. The Facet-Streams interface used in this study included the entire functionality described in the design section, including all numerical feedback and Boolean OR.

Tasks & Procedure. Participants were briefly introduced to the interface for 5 min and had further 5 min to explore it.

Then a single task from the first study was presented as a "warm-up" task. In the following task 1, we presented pre-defined networks of facet tokens on the tabletop to the participants. Each network was manually set up by the experimenter without any explanation. To analyze the participants' understanding of the network, they were handed a printed index card of a hotel with its values for different facets. Participants then had to answer for four different locations in the network whether this hotel was contained in the stream flowing there (Figure 7). Thereby participants were not allowed to alter the network at any time. They were only told whether their answer was correct or wrong without any further explanation. This was repeated with hotels and networks of increasing difficulty. Each participant answered 3 (hotels) x 3 (networks) x 4 (locations in the network) = 36 questions. While instructing participants, we strictly avoided any terminology from Boolean logic or set theory, so that the visualization had to "speak" for itself. Thereby, we hoped to reduce any bias based on prior knowledge of Boolean logic or mathematics.

During task 2, we observed whether participants were able to create complex networks from natural language input. The experimenter played the role of a customer of a travel agency and the participant was asked to use the system to construct a network that answers the customer's questions. The experimenter increased the complexity of his requests in 7 steps following the course of a realistic narration in natural language without referring to concepts from Boolean logic or set theory. To solve this task, participants had to add or remove tokens and to create AND and OR connections between them. Furthermore, they were confronted with conflicting criteria that led to zero results and advanced features had to be used, e.g. multiple output streams from one token. Task 2's complexity at step 7 becomes evident in its Boolean equivalent: (*room quality* = 4-6 OR *location quality* = 5-6) AND (*hotel stars* = 4-5) AND (*country* = Germany OR *country* = Spain) AND (*features* = Bar+Pool). Eventually, participants filled out a user satisfaction questionnaire with semantic-differential adjective pairs and were given a compensation of 7 EUR. Sessions took 30 minutes on average.
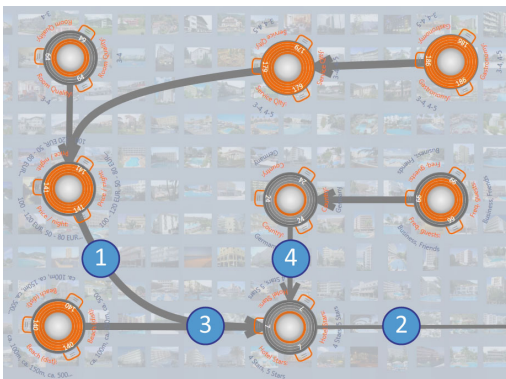

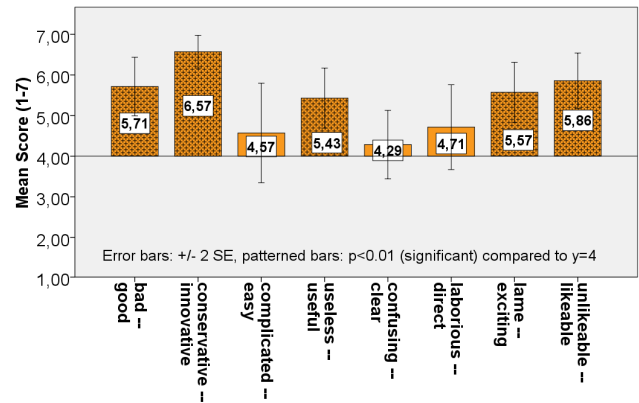**Figure 7: Network from task 1 with question locations ① – ④.**

Figure 8: Semantic differentials from the questionnaires.

Results. For task 1 the total number of incorrect answers was 18 of 252 (7.1%). Regarding the subset of questions that involved a logical OR connection, the number of incorrect answers was 14 of 147 (9.5%). The few incorrect answers show that we succeeded to provide a learnable visual metaphor that conveyed Boolean logic even without user interaction. It was learned by the participants without any interaction or extensive periods of training. This was also true for the more complex networks including AND and OR. During task 2, 6 out of 7 participants managed to correctly construct the entire network including the logical OR from natural language. This is an important insight for co-located collaboration where verbal communication and coordination plays a critical role. Furthermore, task 2 also specifically showed the low viscosity of the interface, as participants frequently lifted and rearranged tokens to try out different alternatives. Figure 8 gives an overview of the results of the questionnaires. The scores support our observation from study 1 that Facet-Streams is perceived as appealing, innovative, exciting, likeable, and useful.

CONCLUSION

We have presented Facet-Streams, a novel design for faceted product search on a tabletop. It proved to be equally effective as established designs for faceted navigation on the Web, although it introduces novel and unfamiliar hybrid interaction techniques and visual metaphors. With respect to our research questions Q1-Q3, we conclude that users perceived using Facet-Streams as a fun experience and considered its design as innovative. An increased awareness and better mutual support among collaborators was observed (Q1). We could also confirm that a great variety of search strategies and collaboration styles can be realized with our design. In particular, we achieved seamless transitions between tightly-coupled collaboration and loosely-coupled parallel work (Q2). Based on the effectiveness during the first study and the small failure rates in the second study, we can confirm that users were able to quickly learn and apply our visual metaphor for Boolean logic. Thereby they also succeeded in formulating complex Boolean queries based on natural language instructions (Q3). Our future work will address a closer integration of query formulation and result browsing to

strengthen our design by enabling a better collaborative sensemaking and evaluation of search results.

REFERENCES

1. Amershi, S., and Morris, M. R. CoSearch: A system for co-located collaborative web search. In *Proc. CHI 2008*, ACM (2008), 1647–1656.

2. Blackwell, A. and Green, T. Notational Systems – The Cognitive Dimensions of Notations Framework. In *HCI Models, Theories and Frameworks*. Morgan Kaufmann, USA, 2003, 103-133.

3. Blackwell, A. F., Stringer, M., Toye, E. F., and Rode, J. A. 2004. Tangible interface for collaborative information retrieval. In *CHI 2004 EA*. ACM (2004).

4. Borgman, C. L. Why are online catalogs still hard to use? *J. Am. Soc. Inf. Sci.* 47, 7 (Jul. 1996), 493-503.

5. Dachselt, R., Frisch, M., Weiland, M.: FacetZoom: a continuous multi-scale widget for navigating hierarchical metadata. In *Proc. CHI 2008,* ACM (2008), 1353-1356.

6. Detken, K., Martinez, C., and Schrader, A. The search wall: tangible information searching for children in public libraries. In *Proc. TEI 2009*. ACM (2009).

7. Elmqvist, N., Stasko, J., Tsigas, P. DataMeadow: A Visual Canvas for Analysis of Large-Scale Multivariate Data. *Information Visualization 7*, 1 (March 2008), 18-33.

8. Hansaki, T., Shizuki, B., Misue, K., and Tanaka, J. FindFlow: visual interface for information search based on intermediate results. In *Proc. APVIS 2006*. ACS (2006), 147-152.

9. Hartmann, B., Morris, M. R., Benko, H., and Wilson, A. Augmenting interactive tables with mice & keyboards. In *Proc. UIST 2009*, ACM (2009), 149-152.

10. Hearst, M. A. *Search User Interfaces.* Cambridge University Press, USA, 2009.

11. Heilig, M., Demarmels, M., Rexhausen, S., Huber, S., Runge, O.: Search, explore and navigate - designing a next generation knowledge media workbench. In: *Proc. SIDeR 2009*, Eindhoven, NL (2009), 40-43.

12. Hornecker, E. and Buur, J. Getting a Grip on Tangible Interaction: A Framework on Physical Space and Social Interaction. In *Proc. of CHI 2006*. ACM (2006).

13. Isenberg, P., and Fisher, D. Collaborative brushing and linking for co-located visual analytics of document collections. In *Proc. EuroViz 2009*, IEEE (2009).

14. Jordà, S., Geiger, G., Alonso, M., and Kaltenbrunner, M. The reacTable: exploring the synergy between live music performance and tabletop tangible interfaces. In *Proc. TEI 2007*. ACM (2007), 139-146.

15. Kirk, D., Sellen, A., Taylor, S., Villar, N., and Izadi, S. 2009. Putting the physical into the digital: issues in designing hybrid interactive surfaces. In *Proc. HCI UK 2009*. BCS (2009), Swinton, UK, 35-44.

16. Lee, B., Smith, G., Robertson, G. G., Czerwinski, M., Tan, D.: FacetLens: exposing trends and relationships to support sensemaking within faceted datasets. In *Proc. CHI 2009*, ACM (2009).

17. Morris, M. R., Paepcke, A., and Winograd, T. TeamSearch: Comparing techniques for co-present collaborative search of digital media. In *Proc. of IEEE Tabletop 2006*, IEEE (2006), 97-104.

18. Morris, M. R., Fisher, D. and Wigdor, D. Search on surfaces: Exploring the potential of interactive tabletops for collaborative search tasks. *Inf. Processing & Management* (2009), Doi:10.1016/j.ipm.2009.10.004.

19. Morris, M. R., Lombardo, J., and Wigdor, D. WeSearch: Supporting collaborative search and sensemaking on a tabletop display. In *Proc. CSCW 2010*, ACM (2010), 401-410.

20. Morville, P., and Callender, J. *Search Patterns - Design for Discovery*. O'Reilly Media, USA, 2010.

21. Norman, D. A. *The Design of Everyday Things*. Basic Books, USA, 2002.

22. Rogers, Y. and Lindley, S. Collaborating Around Large Interactive Displays: Which Way is Best to Meet? *Interacting with Computers*, 2004.

23. Ryall, K., Forlines, C., Shen, C., Morris, M. R., and Everitt, K. Experiences with and Observations of Direct-Touch Tabletops. In *Proc. TABLETOP 2006*. IEEE (2006), 89-96.

24. Shen, C., Lesh, N., Vernier, F., Forlines, C., and Frost, J. Building and sharing digital group histories. In *Proc. CSCW 2002*, ACM (2002). 324-333.

25. Taylor, S., Izadi, S., Kirk, D., Harper, R., and Garcia-Mendoza, A. Turning the tables: an interactive surface for vjing. In *Proc. CHI 2009*. ACM (2009).

26. Tobiasz, M., Isenberg, P., Carpendale, S.: Lark: Coordinating Co-located Collaboration with Information Visualization. *IEEE TVCG* 15, 6 (November 2009), 1065-1072.

27. Ullmer B., Ishii H., and Jacob R. J. K. Tangible Query Interfaces: physically constrained Tokens for manipulating Database Queries. In *INTERACT 2003*.

28. Young, D. and Shneiderman, B. A graphical filter/flow representation of Boolean queries: a prototype implementation and evaluation. *J. Am. Soc. Inf. Sci.* 44, 6 (Jul. 1993), 327-339.