

Persona-ization: Searching on Behalf of Others

Paul Bennett
Microsoft Research
One Microsoft Way
Redmond, WA
pauben@microsoft.com

Alexander Fishkov¹
St. Petersburg Academic University
Academy of Sciences
St. Petersburg, Russia
alexander.fishkov@yandex.ru

Emre Kiciman
Microsoft Research
One Microsoft Way
Redmond, WA
emrek@microsoft.com

ABSTRACT

Many information retrieval tasks involve searching on behalf of others. Example scenarios include searching for a present to give a friend, trying to find “cool” clothes for a teenage child, looking for medical supplies for an elderly relative [1], or planning a group activity that many friends will enjoy. In this paper, we use demographically annotated web search logs to present a large-scale study of such “on behalf of” searches. We develop an exploratory technique for recognizing such searches, and present information to describe and understand the phenomenon, including the demographics of who is searching, who they are searching for and on what topics.

1. INTRODUCTION

Many information retrieval tasks involve searching on behalf of others. Common scenarios include searching for a perfect gift, trying to find “cool” clothes for a teenage child, looking for medical supplies for an elderly relative [1] or planning a group activity that many friends will enjoy. Unlike tasks performed on behalf of oneself, information retrieval tasks performed on behalf of others are prone to two key challenges: First, the searcher is often unfamiliar with the search domain; and secondly, the searcher is often unfamiliar with the mindset of the search *beneficiary*. For example, when searching for a perfect gift for a friend, the searcher might know that the friend loves to cook, but be unfamiliar with the vocabulary used to refer to kitchen gadgets, cookbooks, and ingredients, and moreover, the searcher might not know what kinds of accoutrements the friend might appreciate. The result is an extended exploratory phase of search, where the searcher attempts to better understand both the kinds of results that can be found, as well as the likely mindset of the beneficiary.

In this paper, we present our work on quantitatively characterizing the phenomena of “on-behalf-of” searches. Using demographically identified search query data from comScore, we identify search queries where a searcher of one demographic group is performing a task on behalf of a different demographic group. We do so by focusing our task on queries that include a reference to a demographic group (e.g., [gifts for men over 50yrs]) and identifying instances when the referenced group differs from the searcher’s own demographic group, as identified in the comScore metadata. Our primary goals in this analysis are to characterize the types of relationships between the searcher and beneficiary and identify the most frequent topics where on-behalf-of searches are observed in the logs. A secondary goal is to use these log-driven insights to reflect on how search could be altered to better support on-behalf-of searches.

As for the first question of who, demographically speaking, is searching for whom, we use the search logs to find a reflection of intuitive social patterns in our data. These patterns include patterns consistent with significant-other/marriage relationships (men perform on-behalf-of searches for women of their same age or slightly younger). We find that while some of these relationships are gendered (young men are more likely to search on behalf of men their grandfather’s age; while it is young women who search on behalf of women their grandmother’s age); other relationships are not (e.g. across all age ranges each gender shows a propensity to search on-behalf-of the opposite gender in the same age range as the searcher – quite possibly a spouse or significant other). Moreover, some relationships are symmetric (e.g. mothers are likely to search on-behalf-of their children; and the children are as likely to search on-behalf of their mothers).

Secondly, we characterize the topics and sub-topics for which people search “on-behalf” of others. We find that the most common topics are broadly games, clothing, and, broadly speaking, fashion and taste related. A key commonality of these topics is that individual and demographic preferences vary greatly---e.g., some brands are very popular among young adults, while others are popular among older demographics.

While this analysis method faces challenges, including ambiguities inherent in short, noisy query text, as well as potential systematic biases when searchers do not indicate a beneficiary’s demographics, we believe that the result is still useful as a first large-scale identification and characterization of the phenomenon of on-behalf-of tasks.

Our findings highlight important challenges in supporting on-behalf-of searches. In some sense, it is an extreme form of exploratory search, where a searcher is attempting to understand not only the domain, resolve vocabulary mismatches, etc., but also attempting to learn the beneficiary’s likely preferences for what a “good result” might be.

These observations suggest an opportunity to provide special support for on-behalf-of searches based on the aggregate search behaviors of others in the beneficiary’s demographic groups. We refer to tailoring results to a search beneficiary’s demographic as *persona*-ization – essentially because the searcher can specify a persona which describes the beneficiary who may be a different person than the user but more generally could be used to describe different aspects of the user (e.g. “my business identity” vs. “my social side”). We present a brief discussion of how the aggregate behaviors of searchers within a beneficiary’s demographics might be used to persona-ize the experience for on-behalf-of searchers.

Presented at SIGIR SPS 2015 Workshop, 2015.

¹ Work performed at Microsoft Research.

2. RELATED WORK

Previously, Becker *et al.* [1] studied and reported on the frequency and type of “on-behalf-of” searches in a public library setting in the United States. In particular they report that 63% of users (48.6 million people) use computers in public libraries to seek information or perform a task on behalf of others. Surveys and interviews revealed that this covers a broad range of tasks: from finding and buying car parts for friends to ordering medical supplies for elderly relatives. In the context of public libraries, the searchers tend to be from economically disadvantaged groups that may have limited alternative access points and can search on behalf of others in their community. Furthermore, there was a prevalence for searchers to be younger than the beneficiaries of the information/tasks – indicating that familiarity with technology may play a role. While search within a public library setting on behalf of others is important, we wish to determine what evidence exists to support that users search on behalf of others more generally. Research in the area of collaborative information seeking highlights challenges of explicit or implicit collaboration among two or more people, including through collaborative filtering and recommender systems [8]. While sharing many challenges, including mediation and mitigation of variances in intents and expertise, on-behalf-of search tasks differ in that the beneficiary of a search may not be a party to the search task.

A variety of work has studied both extracting profiles for personalization as well as how to tailor search results given such profiles. A sampling of such work includes extracting short- and long-term topical profiles [7][5][12][10][14], using weighted term-vectors based on long-term desktop activities [13], leveraging a user’s location [2], and using the user’s previous queries and clicks [4][3][11]. However, personalization makes the assumption that the user is searching for themselves with their own preferences in mind. In the setting of persona-ization, this fundamental assumption is challenged, and the need is highlighted for methods that either infer the persona (e.g. from the query) or perform a tailored search where the user is given the option of specifying the persona (e.g. query = [video games] where persona = “8 year old boy who loves Minecraft and Peggles”).

A distinct but related area of research is separating a log from a shared device into streams associated with the distinct users in a house. For example White *et al.* [15] attempt to separate search activity on shared devices and Luo *et al.* [9] seek to differentiate user-viewing of television programs in a household. These studies may be useful in identifying the correct persona to use for tailoring search and recommendation, but in those studies, the user actively using the device is still consuming for themselves. In our setting, we target cases where the user is actively searching on behalf of someone else.

In particular, in contrast to previous work, our focus is on using large-scale search logs to analyze the type and frequency of on-behalf-of search. Furthermore, we provide motivational evidence and a description of one approach to persona-ization that leverages an interface where a user can explicitly specify a persona to conduct a search based on a different user’s demographics.

3. SEARCH ON BEHALF OF OTHERS

Informally, a *search on behalf of others* is when a person is attempting to gather information to help satisfy a need or perceived need of someone other than themselves. For clarity, let us define the *searcher* as the person issuing the queries to the search engine, and the *beneficiary* as the person on whose behalf the searcher is searching.

Gift giving, purchasing or planning on behalf of another (including personal assistants), seeking medical information for another, and planning group activities are all example scenarios where an individual must search for information that will ultimately be used to satisfy someone else’s needs or desires. For the searcher, there are many potential challenges when searching for another. For example, the searcher might not have full information about their target person’s needs or desires. The searcher might not fully understand the given topic or domain, and may not even know the necessary vocabulary.

Our characterization of on-behalf-of demographics relies on identifying queries that specify a persona, extracting the topic of the task, identifying the demographics of the searchers from user information sources and extracting the demographics of the beneficiaries from the queries.

3.1 Identifying Queries Specifying a Persona

In this section, we describe a simple way to identify queries that explicitly specify a persona. We use a sample of one month of Bing query logs from May 2013. To reduce linguistic variation due to language and geography, we use only queries from the English US market. From these logs, we extract queries that contain one of the phrases “for women”, “for men”, “for boys”, or “for girls” and then remove any query whose primary topic is identified as adult, pornography, or romance related. We consider the remainder to be queries that specify a persona. These contain such queries as [birthday party ideas for boys in their teens] and [math games for girls]. Overall, we find the frequency of such explicitly-specified queries to be 0.10% of all queries. Note that this is likely an underestimate of the number of both searches that have a persona in mind as well as searches on behalf since it requires matches to a fairly strict phrasing. Search interfaces that enable directly entering and leveraging a user-specified persona, likely would see a greater frequency of use. Nonetheless, given the large number of matches to even this strict phrasing, we focus on analyzing this subset as an initial foray into this domain.

Table 1: Top 20 queries that frequently contain gender/age targets

Topics	Relative % of Persona Queries
games	22.00%
shoes	10.07%
hairstyles	6.42%
tattoos	4.48%
dresses	3.74%
clothing	2.60%
shirts	2.41%
names	2.14%
sandals	2.00%
watches	1.96%
boots	1.95%
gifts	1.83%
haircuts	1.66%
clothes	1.53%
ideas	1.50%
pants	1.44%
swimsuits	1.44%
hats	1.25%
shorts	1.20%
suits	1.18%

In later sections, we will use demographics data to identify whether the persona-specified does or does not match the searcher. However, here we seek to answer the question: what topics searchers are looking for when they explicitly specify a persona?

Table 1 shows the top topics that frequently contain a persona specification. We identify topic by simply extracting the word that precedes the persona specifier (the phrases used above like “for men”). A more general future direction is not only more robustly identifying persona specifications but identifying and normalizing for topic in a more general fashion. Using this simple identification of topic, we find that the most common topics are broadly games, clothing, and, broadly speaking, fashion and taste related. A key commonality of such topics is that individual and demographic preferences vary greatly within these domains---e.g., some brands are very popular among young adults, while others are popular among older demographics. Presumably searchers believe that by explicitly specifying the demographics of the beneficiary, they are more likely to receive appropriately tailored search results.

3.2 Data and Searcher Demographics

To separate searches where the persona refers to the searcher versus a beneficiary other than the searcher, we require both the searcher’s demographics and the beneficiary’s demographics. To this end, we use search logs available from the internet analytics company comScore (comScore.com) with a paid subscription. We use comScore data gathered during January to February 2014 in the English-speaking segment of the US market.

Households opt-in to provide traces of their online activities, including searches at major search engines. Additionally, each person has a unique person identifier that should be used to sign-in before a user in the household uses a particular device for search. ComScore intentionally tries to construct an overall sample that is representative in the target market [6] across all major search providers. Users provide individual demographic data associated with the person identifier. This information provides us with both queries and demographic information about the searcher. Individuals are aware, of course, of comScore’s data collection and are compensated in exchange.

3.3 Beneficiary Demographics

When searching for others, people often embed specific demographic attributes in their query formulations. Knowing the demographic attributes of the searcher, we can recognize when they are searching for another by looking for demographic attributes that do not match their own. For example:

- [shoes] vs. [shoes for women]
- [workouts] vs. [workouts for men over 50]

To infer demographic information about the beneficiary of a query, we extract the gender and age from query text.

In order to accomplish this, we use a small set of rules similar to those in Section 3.1. However, we broaden the set for gender to not necessarily require the leading word “for” as well as include several other gendered expressions, e.g. (“dad”, “mother”). For age, we similarly recognize a small number of patterns denoting age (e.g. “NUMBER yo”, “over NUMBER”, etc.). Challenges are primarily due to recognizing and filtering ambiguous uses of numbers when they are not age-related, (“shoes for women under 20” often refers to the price of the shoe, not the age; “dresses for women under 100” refers to weight or price). Although our patterns recognize cases where a trailing word disambiguates (e.g. we recognized that “shoes for women over 20 dollars” is a price). Again more general

query parsing to extract age and gender is an interesting area for future study. We focus on a narrow set that can be extracted relatively confidently.

If both the age and gender of the beneficiary can be extracted from the query, we assume the beneficiary is the same as the searcher if they have a non-empty intersection (e.g. a 54 year-old man searching for [workouts for men over 50] is assumed to be searching for himself). If we cannot accurately extract both age and gender from the query, then we assume the searcher is searching for themselves to take a conservative analysis point with respect to the phenomenon of on-behalf-of searches. Extending the dataset under consideration (e.g., through session-level or task-level identification of beneficiary demographics; or through other inference methods) remains future work.

4. CHARACTERIZING ON BEHALF OF SEARCHES

To better understand on-behalf-of searches, we wish to characterize the relationship between searchers and beneficiaries as well as understand case studies of breakdowns on some of the most common topics of on-behalf-of searches that were discussed in Section 3.1.

4.1 Demographics of “on behalf of searches”

Table 2 and Table 3 show who is searching for whom. The gender and age of the searcher are shown on the x-axis, and for Table 3, the gender and age of the beneficiary, extracted from the query, are shown on the y-axis. For Table 2, the first column is the percentage relative to any persona-specified query that is issued by that demographic (i.e. the column sums to 100%). The remaining cells are the breakdown within the searcher demographic (i.e. the second and third percentage in a row sums to 100%).

We see that when users specify a persona, they most commonly specify a persona that does not match their own demographics – with nearly all searcher demographics searching on-behalf-of others more than half of the time, with males searching for others more frequently across comparable ages, as compared to females.

		% Persona-Specified	Self	Others
Female	18-20	8.27%	36%	64%
	21-24	9.34%	6%	94%
	25-34	10.07%	31%	69%
	35-44	8.08%	30%	70%
	45-49	3.40%	39%	61%
	50-54	2.94%	20%	80%
	55-64	3.44%	53%	47%
	65+	1.54%	24%	76%
Male	18-20	8.54%	11%	89%
	21-24	12.76%	9%	91%
	25-34	13.33%	7%	93%
	35-44	6.63%	22%	78%
	45-49	3.27%	8%	92%
	50-54	2.81%	15%	85%
	55-64	3.42%	10%	90%
	65+	2.17%	2%	98%

Table 2: Percentage of all queries specifying a persona by searcher demographic and breakdown into self-search and on-behalf search.

		Beneficiary																		
		Female								Male										
		18-20	21-24	25-34	35-44	45-49	50-54	55-64	65+	18-20	21-24	25-34	35-44	45-49	50-54	55-64	65+			
Searcher	Female	18-20	36%	2%	9%	5%	7%	6%	5%	0%	18-20	10%	5%	4%	2%	1%	4%	3%	1%	18-20
	21-24	15%	6%	10%	8%	22%	1%	14%	5%	21-24	6%	3%	1%	5%	0%	0%	3%	1%	21-24	
	25-34	5%	4%	31%	11%	4%	12%	9%	4%	25-34	2%	1%	4%	1%	2%	2%	4%	4%	25-34	
	35-44	10%	3%	3%	30%	13%	0%	5%	7%	35-44	4%	0%	1%	9%	6%	3%	4%	1%	35-44	
	45-49	3%	0%	1%	13%	39%	9%	9%	0%	45-49	12%	0%	0%	4%	4%	2%	3%	2%	45-49	
	50-54	2%	0%	27%	3%	4%	20%	28%	3%	50-54	2%	0%	0%	0%	4%	1%	2%	3%	50-54	
	55-64	1%	0%	3%	3%	2%	10%	53%	17%	55-64	0%	0%	1%	1%	0%	3%	6%	1%	55-64	
65+	0%	0%	0%	2%	8%	24%	35%	24%	65+	0%	0%	0%	1%	0%	0%	2%	2%	65+		
Searcher	Male	18-20	34%	4%	4%	9%	7%	8%	3%	1%	18-20	11%	0%	4%	5%	6%	0%	2%	1%	18-20
	21-24	26%	4%	4%	7%	3%	6%	5%	2%	21-24	9%	9%	4%	4%	2%	0%	15%	0%	21-24	
	25-34	19%	5%	9%	9%	9%	5%	11%	5%	25-34	8%	1%	7%	4%	1%	3%	4%	1%	25-34	
	35-44	17%	1%	0%	14%	8%	2%	10%	8%	35-44	2%	0%	7%	22%	8%	0%	1%	1%	35-44	
	45-49	16%	1%	6%	16%	20%	1%	4%	12%	45-49	0%	0%	0%	8%	8%	5%	0%	5%	45-49	
	50-54	16%	1%	12%	7%	22%	7%	11%	0%	50-54	2%	1%	0%	0%	2%	15%	3%	1%	50-54	
	55-64	15%	0%	6%	14%	13%	8%	21%	6%	55-64	1%	2%	0%	0%	0%	3%	10%	1%	55-64	
65+	4%	0%	2%	10%	14%	4%	14%	32%	65+	3%	0%	1%	1%	3%	10%	1%	2%	65+		

Table 3: Demographics of beneficiary for specified persona in query relative to searcher's demographics

		Beneficiary																		
		Female								Male										
		18-20	21-24	25-34	35-44	45-49	50-54	55-64	65+	18-20	21-24	25-34	35-44	45-49	50-54	55-64	65+			
Searcher	Female	18-20	73%	4%	0%	0%	0%	0%	5%	0%	18-20	18%	0%	0%	0%	0%	0%	0%	0%	18-20
	21-24	52%	0%	0%	0%	16%	0%	30%	0%	21-24	2%	0%	0%	0%	0%	0%	0%	0%	21-24	
	25-34	13%	0%	52%	0%	0%	9%	0%	17%	25-34	0%	9%	0%	0%	0%	0%	0%	0%	25-34	
	35-44	0%	0%	0%	0%	20%	0%	12%	0%	35-44	68%	0%	0%	0%	0%	0%	0%	0%	35-44	
	45-49	0%	0%	0%	22%	42%	0%	32%	0%	45-49	4%	0%	0%	0%	0%	0%	0%	0%	45-49	
	50-54	2%	0%	79%	1%	5%	5%	8%	0%	50-54	0%	0%	0%	0%	0%	0%	0%	0%	50-54	
	55-64	0%	0%	0%	12%	0%	0%	62%	26%	55-64	0%	0%	0%	0%	0%	0%	0%	0%	55-64	
65+	0%	0%	0%	0%	0%	0%	18%	82%	65+	0%	0%	0%	0%	0%	0%	0%	0%	65+		
Searcher	Male	18-20	24%	0%	20%	0%	0%	0%	0%	0%	18-20	13%	0%	0%	2%	17%	0%	24%	0%	18-20
	21-24	8%	0%	20%	0%	13%	0%	0%	0%	21-24	9%	0%	0%	5%	42%	0%	2%	0%	21-24	
	25-34	4%	0%	0%	20%	52%	0%	0%	0%	25-34	18%	0%	0%	0%	6%	0%	0%	0%	25-34	
	35-44	0%	0%	0%	42%	0%	0%	15%	0%	35-44	0%	0%	0%	8%	0%	0%	34%	0%	35-44	
	45-49	0%	0%	10%	0%	46%	0%	0%	44%	45-49	0%	0%	0%	0%	0%	0%	0%	0%	45-49	
	50-54									50-54									50-54	
	55-64	0%	0%	0%	0%	0%	3%	0%	0%	55-64	0%	0%	0%	0%	0%	0%	97%	0%	55-64	
65+	0%	0%	0%	0%	0%	0%	0%	100%	65+	0%	0%	0%	0%	0%	0%	0%	0%	65+		

Table 4: In "clothes" category, demographics of beneficiary relative to searcher's demographics

From the detailed breakdown in Table 3, we see several patterns of searches reflecting intuitive relationships among demographic groups:

- People likely searching on-behalf-of their children (e.g. 27% of persona-specified queries for women ages 50-54 have a beneficiary of women ages 25-34). E.g., [christmas gifts for a 25 year old woman]
- People likely searching on-behalf-of their parents and grandparents (e.g. 15% of persona-specified queries for men ages 21-24 have a beneficiary of men ages 55-64; 5% of persona-specified queries for women ages 21-24 have a beneficiary of women age 65+). E.g., [present for grandma 60 years]
- People searching on-behalf-of their spouses/significant others: note the significant diagonal in the *opposite* gender from the searcher in the same age category. E.g., [birthday gifts 30 year old female]

4.2 Demographics and Topics of “on behalf of searches”

We can also consider certain topical slices. Table 4 presents a similar table to Table 3, but restricted to queries that mention clothes. Note that some cells are missing here due to sparsity given the sample size.

Looking at the relationship between demographics and topics, we see that many of the symmetric relationships we identified in Section 4.1 seem to lose their symmetry. For example, while mothers look for clothes on behalf of their children, their children do not necessarily search for clothes on behalf of their parents.

5. DISCUSSION

5.1 Using Self-Searches to Aid On-Behalf-Of Searches

In Table 5, we can see how style- and brand-popularity varies significantly across gender and age groups. Only a small number of keywords (Jordan and Nike) are popular across most age groups and both genders. Some (for example, Vans, Converse) are popular only among a single demographic. Other observations include that at all comparable ages, a larger proportion of males’

	Searcher									Searcher								
	Female									Male								
	18-20	21-24	25-34	35-44	45-49	50-54	55-64	65+		18-20	21-24	25-34	35-44	45-49	50-54	55-64	65+	
adidas	4%	4%	3%	3%	2%	3%	2%	4%	6%	4%	4%	5%	3%	4%	2%	2%		
bakers	2%	3%	3%	2%	2%	2%	2%	1%	1%	1%	1%	1%	1%	1%	1%	1%		
basketball	5%	5%	4%	7%	8%	4%	5%	3%	9%	11%	8%	10%	10%	7%	5%	8%		
boat	2%	2%	2%	2%	1%	2%	1%	2%	3%	3%	4%	1%	2%	3%	3%	1%		
cheap	3%	4%	5%	2%	3%	1%	1%	0%	2%	2%	3%	2%	1%	2%	1%	1%		
converse	3%	3%	3%	3%	2%	2%	1%	2%	2%	2%	2%	2%	3%	2%	1%	1%		
dance	2%	2%	2%	2%	2%	3%	4%	5%	1%	0%	1%	1%	1%	1%	1%	5%		
dc	2%	2%	2%	2%	2%	3%	1%	1%	2%	3%	2%	2%	1%	1%	1%	0%		
dress	3%	4%	3%	4%	4%	6%	8%	10%	3%	3%	4%	5%	6%	5%	5%	7%		
golf	0%	0%	1%	1%	1%	1%	2%	6%	1%	1%	3%	3%	3%	6%	8%	13%		
gucci	1%	1%	1%	1%	1%	1%	1%	1%	2%	2%	2%	2%	1%	1%	0%	0%		
heel	3%	3%	5%	2%	2%	1%	2%	3%	0%	1%	1%	1%	1%	1%	1%	1%		
james	2%	2%	2%	2%	2%	1%	1%	0%	3%	3%	2%	3%	3%	1%	1%	1%		
jordan	14%	12%	10%	10%	9%	8%	8%	4%	18%	15%	14%	11%	7%	8%	7%	7%		
jordans	3%	3%	2%	2%	2%	1%	1%	1%	3%	3%	2%	2%	1%	1%	1%	1%		
mens	1%	1%	2%	3%	2%	2%	2%	4%	2%	2%	3%	2%	3%	4%	6%	5%		
nike	9%	8%	7%	7%	6%	6%	4%	2%	10%	8%	9%	9%	10%	5%	4%	5%		
payless	6%	7%	10%	10%	14%	19%	17%	15%	2%	2%	3%	6%	11%	10%	9%	8%		
room	3%	3%	3%	4%	3%	3%	4%	4%	1%	1%	1%	2%	2%	4%	3%	3%		
running	6%	6%	6%	7%	9%	7%	5%	5%	7%	8%	11%	9%	10%	13%	17%	11%		
skate	2%	3%	1%	2%	2%	2%	1%	1%	3%	4%	3%	2%	2%	2%	1%	1%		
soccer	3%	2%	3%	2%	1%	1%	1%	1%	5%	5%	4%	4%	5%	2%	1%	1%		
supra	2%	1%	1%	2%	2%	1%	0%	0%	4%	4%	2%	3%	1%	2%	2%	3%		
tennis	4%	4%	4%	5%	6%	6%	7%	9%	2%	2%	4%	5%	4%	5%	10%	3%		
vans	3%	3%	1%	2%	1%	2%	2%	1%	3%	3%	2%	1%	2%	1%	1%	2%		
wedding	4%	4%	6%	2%	3%	2%	3%	2%	1%	1%	2%	1%	1%	2%	1%	1%		
women	2%	3%	4%	2%	2%	3%	4%	5%	0%	0%	0%	0%	0%	0%	0%	0%		
womens	2%	2%	3%	2%	3%	4%	6%	7%	1%	1%	1%	1%	2%	2%	3%	5%		
wrestling	1%	0%	1%	2%	2%	2%	0%	1%	2%	2%	2%	3%	2%	2%	1%	1%		
your	2%	1%	1%	2%	1%	2%	1%	1%	2%	3%	2%	1%	2%	2%	2%	1%		

Table 5: Relative popularity of keywords used, $P(\text{keyword} | \text{age, gender})$, when searching for “shoes”, by demographic.

shoe queries are related to golf and running shoes as compared to females. Likewise, among males, the proportion of shoe queries related to basketball, Jordan, Nike and Adidas shoes decreases with increasing age, while queries related to dress, golf, and running shoes increase with age. These types of trends are not obvious to any searcher who wishes to search on behalf-of-another demographic. By identifying the prevalent particular items within a targeted category, search data can be used to improve a searcher's exploration by exposing to them – either indirectly through re-ranked results or directly through suggested queries/items – the behavior of the demographic they wish to search on-behalf-of.

5.2 Beyond Demographics

When people search on-behalf-of another, they are not limited to characterizing a beneficiary based on the demographic details. Searchers often know other key information, such as their likes and preferences and embed this information into queries ([games for people who like star wars]). In the same way that we used age and gender information embedded in a query to identify beneficiary demographics, we can use these hints to identify a cohort similar to a given beneficiary. That is, we can identify searchers from the targeted demographic whose search history indicates a preference from star wars. That identified cohort can then be restricted to game-related queries and the resulting queries can be used to expose to the searcher games that would likely match the interest and demographics of the beneficiary.

6. SUMMARY

In this short paper, we provide the first large-scale characterization of the phenomena of on-behalf-of searches. We characterized the demographic relationships embedded within such searches, as well as the primary topics that are the subject of such searches, highlighting key challenges faced by on-behalf-of searchers when attempting to complete their tasks.

In future work, we would like to explore on-behalf-of searches at a session level, looking at metrics of satisfaction and task completion, and further exploring how the aggregate behaviors of beneficiary cohorts can be used to improve the on-behalf-of search experience.

7. REFERENCES

[1] S., Becker, M.D. Crandall, K.E. Fisher, B. Kinney, C. Landry, and A. Rocha. Opportunity for All: How the American Public Benefits from Internet Access at U.S. Libraries. (IMLS-2010-RES-01). Institute of Museum and Library Services. Washington, D.C. 2010.

- http://impact.ischool.washington.edu/documents/OPP4ALL_FinalReport.pdf
- [2] P.N. Bennett, F. Radlinski, R.W. White, and E. Yilmaz. Inferring and Using Location Metadata to Personalize Web Search. In SIGIR '11, pp. 135–144, 2011.
- [3] H. Cao, D.H. Hu, D. Shen, D. Jiang, J. Sun, E. Chen, and Q. Yang. Context-aware Query Classification. In SIGIR '09, pp. 3–10, 2009.
- [4] H. Cao, D. Jiang, J. Pei, Q. He, Z. Liao, E. Chen, and H. Li. Context-aware query suggestion by mining click-through and session data. In KDD '08, pp. 875–883, 2008.
- [5] P.A. Chirita, W. Nejdl, R. Paiu, and C. Kohlschütter. Using ODP Metadata to Personalize Search. In SIGIR '05, pp. 178–185, 2005.
- [6] Fulgoni, G.M. The "Professional Respondent" Problem in Online Survey Panels Today. Slides online at: http://www.sigvalidation.com/tips/05_06_02_Online_Survey_Panels.ppt (Downloaded on June 1, 2015). 2005.
- [7] S. Gauch, J. Chaffee, and A. Pretschner. Ontology-Based User Profiles for Search and Browsing. *Web Intelligence and Agent Systems 1*, 3–4, pp. 219–234, 2004.
- [8] G. Golovchinsky, P. Qvarfordt, J. Pickens. Collaborative Information Seeking. *IEEE Computer*, 42(3): 47–51, 2009.
- [9] D. Luo, H. Xu, Hongyuan Zha, J. Du, R. Xie, X. Yang, and W. Zhang. You Are What You Watch and When You Watch: Inferring Household Structures from IPTV Viewing Data. *IEEE Transactions on Broadcasting Technology*, 60(1): 61–72, 2014.
- [10] Z. Ma, G. Pant, and O.R.L. Sheng. Interest-based Personalized Search. *TOIS* (25:1), Article 5, 2007.
- [11] L. Mihalkova and R. Mooney. Learning to Disambiguate Search Queries from Short Sessions. In *ECML PKDD '09*, pp. 111–127, 2009.
- [12] M. Speretta and S. Gauch. Personalized search based on user search histories. In *Web Intelligence '05*, pp. 622–628, 2005.
- [13] J. Teevan, S.T. Dumais, and E. Horvitz. Personalizing Search via Automated Analysis of Interests and Activities. In *SIGIR '05*, pp. 449–456, 2005.
- [14] R.W. White, P.N. Bennett, and S.T. Dumais. Predicting short-term interests using activity-based search context. In *CIKM '10*, pp. 1009–1018, 2010.
- [15] R.W. White, A. Hassan, A. Singla, and E. Horvitz. From Devices to People: Attribution of Search Activity in Multi-User Settings. In *WWW '14*, pp. 431–442, 2014.