User Modeling in search for People with Autism

Author: **Esha Massand**, Supervisor: **Keith Mannock**

Birkbeck, University of London
Department of Computer Science and Information Systems

Project report submitted in partial fulfillment of the requirement for the
MSc in Computer Science

August 2015

Abstract

This project report presents the development of a prototype web application to assist users with Autism when they search the web. The developed system models user interactions with the search process into a user profile by integrating insights from the core features of autism into the model. The user profile is applied to a synthesis of three leading search engines, and the entire system is integrated with an infra-red user interface component to assist users with Autism during search.

word count (proposal text only) : 3294 words.

Contents

Abbreviations

API	Application Programming Interface
AQ	Autism Quotient
ASD	Autism Spectrum Disorder
DSM	Diagnostic and Statistical Manual
GCS	Google Custom Search
HCI	Human Computer Interaction
HTTP	Hypertext Transfer Protocol
JSON	JavaScript Object Notation
IDE	Integrated Development Environment
KWIC	Key Word In Context
LEAP	LEAP Motion Controller
REST	Representational state transfer
RIFT	Oculus Rift Virtual Reality Head Mounted Display
TDD	Test Driven Development
UI	User Interface
UX	User Experience
VR	Virtual Reality

Definitions

Autism Autism is amongst the most common neurodevelopmental condition and it is currently estimated that 1/68 children meet criteria for Autism Spectrum [**?**]. Autism is five times more common amongst boys than girls (1/42 boys, and 1/189 girls). According to the Diagnostic and Statistical Manual (2013), Autism is characterized by persistent and early deficits in reciprocal social interaction and repetitive behaviours. Individuals vary from high functioning to low functioning (along a spectrum), with behaviours emerging around 2 to 3 years of age.

1 Introduction

This project report presents the development and evaluation of a prototype web-browser based application to assist users with Autism when they search the web, hereafter referred to as Jellibeans [1]. Jellibeans will run in a web browser and utilises gesture and hand movement data recorded using the Leap Motion Controller (LEAP).

Individual characteristics of each user will be measured using a 50 item questionnaire, the Autism Quotient (AQ [1]) which measures tendency towards autistic traits. A score of 32 or higher indicates a strong likelihood of Autism or Asperger syndrome (the questionnaire has a 79% sensitivity score). Individuals who score highly on the AQ will be offered the current user model for their search. The questionnaire data will be stored in the user's Google+ profile????

In programmatic terms, Jellibeans is designed to implement a research-based user model in search. This itself is a considerable part of the current research project and involves the development of a rule engine to transform the idiosyncratic ??? nature of search queries formed by individuals with ASD to work with current available search engines. This development involved collecting and analysing search behaviour patterns from people with and without ASD and building a set of transformation rules given the features of search queries that are formed by individuals with ASD.

Jellibeans will also improve the search experience for users with ASD by going one step further and enabling motion controlled search using the LEAP motion controller. The interface will be dynamic as opposed to static and this afford many advantages over traditional search engine interfaces.

1.1 User Models in Search

1.2 Rule and Transformation Engine

1.3 Motion Controlled Navigation in Search

2 Aims

The goal of the project is to build a prototype search tool that assists users with Autism search and navigate the web. To acheive this goal, the aims of this work are:

To synthesise the search results from three search engines, Google, Bing and Yahoo. For search results returned by Google, the Custom Search API will be used in line with the Google terms of service (as 'screen scraping', or copying the data directly from the website is prohibited). It is a RESTful api with a single method called list. The API method used was GET, and the response data is returned as a JSON type. The response consists of (1) the actual search result, (2) metadata for search like number of results, alternative search queries, and (3) custom search engine metadata. The data model depends on [5]. For Bing and Yahoo search results, JSoup (a Java HTML parser) will be used to identify the links from the resulting query. The JSoup HTML parser was considered more effient for retrieving search results, as it reduced the number of lines of code required to complete the task. Jsoup has its advantages over html parsing. It contains a class representing a list of nodes, 'Elements', which implements Iterable to iterate over a list in an enhanced for loop. The resulting links are written to text file which stores the links in a text file in the projects source directory.

The system aims to deliver relevant results to the user, but what is "relevant"? Not all users or groups of users search in the same way, so it is important to consider user intention. The first

---

[1]Jellibeans are a rainbow of colours, different sizes and shades, and the name represents the difference in style of processing of individuals with ASD.

part of the current project aimed to investigate the differences between search queries of people with and without Autism, and to use those findings to build a stereotyped user model (a user model that will infer characteristics about the user from their diagnostic information) for a person with autism. Users will have to register will Google+ ???, and report their diagnostic information in the aboutMe section of their profile ??? / complete the Autism Quotient Questionnaire and obtain a score of autistic-like traits. Using the Google+ API, Jellibeans will connect to Google+ and ask the user to signin/ agree to the web application accessing their data. JavaScript will be used to parse the text in the aboutMe section to check if the user has identified as a persn with autism, aspergers, ASD or not. The result is returned via the console (inspect) command. ???

3 Method

3.1 Building a User Model into Search

Search queries usually fall into one of three broad categories. 'Do' queries which characterise transactions between the user and the search engine, for example when the user wants to do something such as buy a plane ticket or listen to a song. There are also 'Know' type queries. These are informational queries for example, the name of a band or restaurant in London. The third broad category is 'Go' type queries, which are navigational in nature, for example, searching for a particular home page on the web.

There are many stages to the search process. After identifying the information need, the user must formulate a search query. The user must browse through results once the query has been entered into a search engine. The whole process can be repeated if the user is not satisfied. The stage at which the user model will have most possible impact is before results are returned to the user, that is, at the query formulation stage. It is this stage that the user model will be implemeted (stage 2 on Figure 3.1).

1. Experience the need for an answer, solution, or piece of information.

2. Formulate that need in a string of words and phrases, also known as the query.

3. Enter the query into a search engine.

4. Browse through the results for a match.

5. Click on a result

6. Scan for a solution, or a link to that solution.

7. If unsatisfied, return to the search results and browse for another link or ...

8. Perform a new search with refinements to the query
<div align="center">Typical user's search flow process [6]</div>

3.2 Data Collection

To identify the features of the user model to build, I ran a study to collect example search queries on a set of informational needs from 33 participants. The participants were asked to give examples of search queries they would use to identify the name of a song they had heard (given the lyrics), or the name of a breed of a dog they had seen (given a picture of the dog). There were in total 10 search queries; the study was distrbuted widely via Surveymonkey.com [7] and can be seen in Appendix 7.1. Participants were also asked to complete the Autism Spectrum Quotient

50-item questionnaire see Appendix 7.2. The participants responses are analysed and reported back to surveymonkey.

The range of scores for the AQ is 0 to 50 with high scores indicating increased liklihood of autism-like traits. A score under 21 is a low to average result (many women average around 15 and men around 17). A score of 22-25 indicates autistic tendencies slightly above the population average. A score above 26 gives a borderline indication of high functioning autism, or aspergers. A score above 32 suggests a likelihood of Asperger syndrome or autism (sensitivity of test measure = 79% [1]). However although the AQ is a useful tool, it is by no means a form of diagnosis of autism. For the purposes of this study, individuals with scores equal to, and above 30 were interpreted as having 'autistic-like traits'.

Participants were divided into two groups; low AQ scorers (scores below 30), and high AQ scorers (scores equal to and above 30). There were 29 low AQ scorers and 7 high AQ scorers.

3.3 Differences in Search Queries Between Users With and Without Autistic-like traits.

I conducted a qualitative analysis on the search query strings from both low and high AQ scorer groups.

In both groups, Google was the prefered search engine by far, with all participants reporting that they used Google as a first choice. No one in the current sample used Yahoo or Bing. The low AQ scorer responses were analysed as a group. A 'model' (or baseline) was generated using a frequency criterion of 40% i.e., if 12 out of 29 respondents or more generated the same query string given an informational need, it was included in the model below. If two responses were equally as common, both are reported in the model. When a response indicated that the participant would do an image search, I was unable use those data. The results from the frequency analysis are presented below.

1. You hear a song on the radio with the lyrics, Look at your children, and you want to download it. What would you type into search on your favourite search engine to find out what song it was?
   Look at your children song.
   Look at your children lyrics.

2. Youve lost touch with an old school friend (you went to St. Marys School). What key words/queries would you use to find them?
   St. Marys School Year of X.

3. How would you identify what this is using a search engine (pretend you dont know what it is called). What key words/queries would you use?
   Star shaped brown plant.

4. How would you find out the name of this famous person using a search engine? What key words/queries would you use?
   Brown hair famous young women.

5. How would you identify what breed this animal is using a search engine? What key words/queries would you use?
   Small dog fluffy breed.

6. Your friend and you cant agree on how Thandie Newton pronounces her first name. How would you resolve this using a search engine?
   Thandie Newton pronunciation.

7. What would you search for to identify this patterns name, and which country it originates from?
   Repeating square maze pattern border.

8. How would you search for delays relating to your (imminent) flight to Paris?
   Flight number, carrier, Paris, airport, flight time.

The following observations were made for the users in the high AQ group:

1. Although high AQ scorers' for the most part, formed query strings accurately, there was increased use of idiosyncratic words in the query strings that were forumulated. This is in line with previous research that suggests that people with autism organise information in more subjective and individual ways [2]. For example, referring to the picture of the dog as 'yorkie pooh'(not listed in frequency index), 'aeroplane' (frequency of 8254 words per million) instead of plane (frequency of 33900 words per million) or flight (29535 words per million) , 'miniature' (less than 4973 words per million) instead of small (185463 words per million). The idiosyncratic nature of the words is captured by their lower frequency of use in the English language. Search engines use term frequency to determine if a document is relevant to the users search query. If the frequency of words used to form search queries differs between low and high AQ scorers, so will the returned search results.

2. There were an increased number of imcompletely-formed queries, in other words, in the high AQ group, participants were more likely to miss off words in the query string. This was observed even though the sample size was much smaller in the high AQ group (7 people) compared to the low AQ group (29 people)). For example, when analysing the resuls from query 1 above, 2 out of 7 respondents in the high AQ group did not put 'lyrics' or 'song' in the search query when searching for the lyrics "Look at your children". When these search strings are entered into Google, the results are very different (see Figure 1), suggesting that Google is not performing well for the high AQ user group.
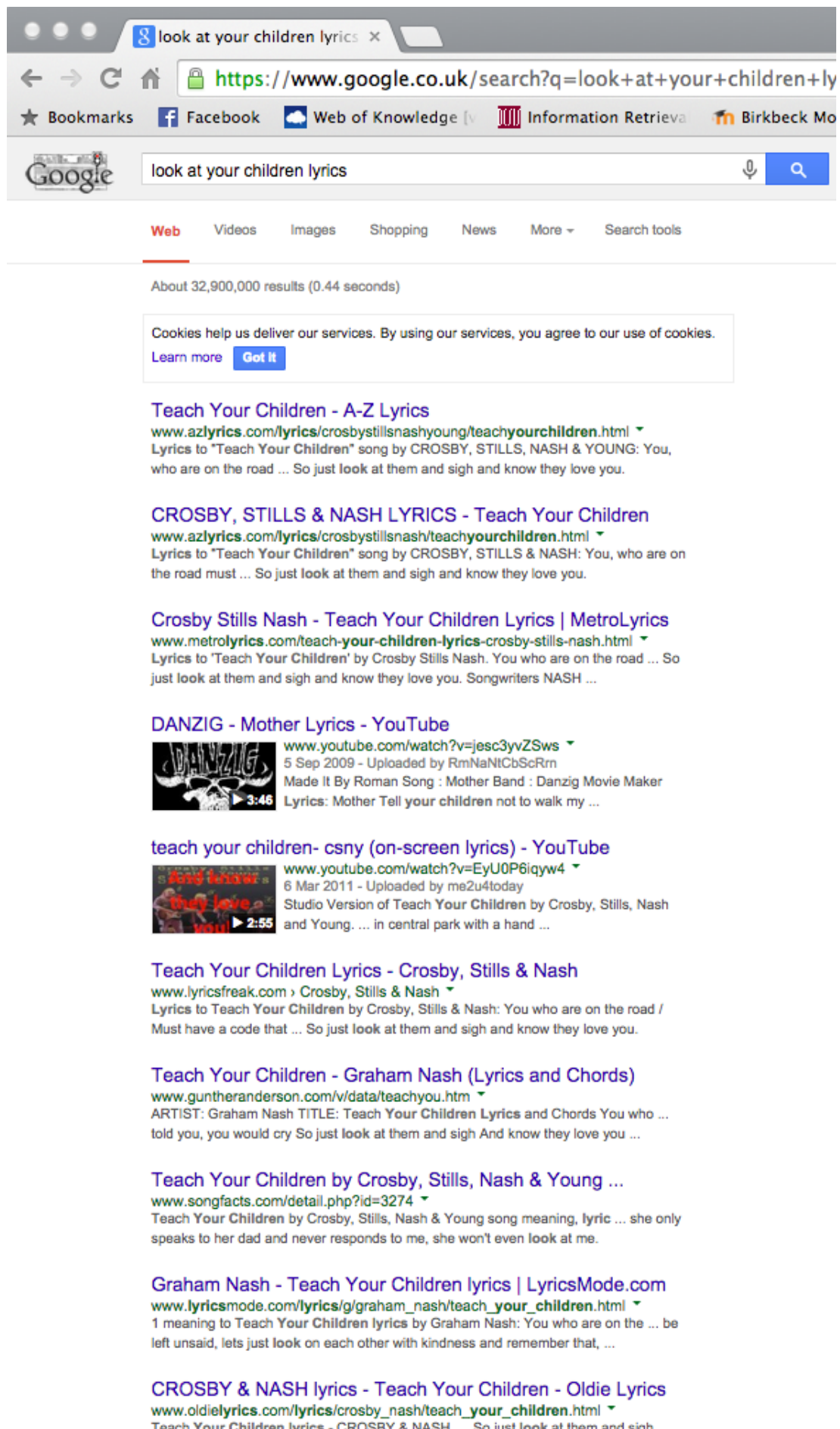
Figure 1: High and low AQ scorers both formed search queries accurately, however there was an increased tendency to omit the word "lyrics" in the high AQ group resulting in very different search results.

3. One individual of the 7 individuals in the high AQ group demonstrated ambiguous use of third-person pronouns, which is characteristic of some individuals with Autism [4]. This includes using first names instead of 'I' or 'you'. This is particularly detrimental to search engine query strings because the use of names distrupts the term frequency - inverse document frequency weighting [8] of the search query and subsequently the results returned to the user.

4. For questions that were 'social' in nature (e.g., featuring a face of a famous woman), 2 out of 7 individuals in the high AQ group indicated that these were types of queries that they would not normally be interested in, and so "wouldn't bother asking it". This is, of course, in line with the characteristics of Autism according to [3]. For these queries, it was more common for individuals in the high AQ group to include information in their search query string that was extraneous to the search question itself, compared to the low AQ group. For example, in query 4 above (which asked respondents to indicate how they would identify a famous person), 2 high AQ scorers included information about the woman's earing. Inevitably this 'dilutes' the search query and results in reduced precision for the search engine.

5. Lastly there were more spelling errors and typographical errors in the high AQ group compared to the low AQ group.

3.4   Building the Jellibeans Transformation Rule Engine

Given the set of observations in the data reported above, we aimed to devise a set of rules that would 'transform' queries made by individuals in the high AQ group to those similar to the low AQ group. The search engines already handle some of the observations from high AQ scorer queries. For example, the use of pronouns is already taken care of with the use of stop words, which is implemented by the search engine itself. The aim of the project is to therefore address issues that result in the search query string being misleading, and returning differnt results to low AQ scorer search queries. Therefore, the set of rules developed (the 'rule engine'), when applied to high AQ user queries, will return search results that are more in line with low AQ scorers. In other words, search results from high AQ queries shared greater overlap with those made by the low AQ group, even though the queries were formed differently. The rule engine is thus a conrete and operationalised framework, for a theortetically-grounded user model of autism within search.

Some of the rules in the rule engine work to address more than one observation made during the data collection stage. As a general rule, changes will take the form of 'suggestions', so that at each stage the user can decide whether to take, or drop the suggestion. I will now introduce each rule and how they were operationalised.

To address observation 1 and to make high AQ scorer search queries less idiosyncratic, words not in english dictionary will be suggested for removal. For example, after each non-word, the user will be prompted with, "This word is not in the dictionary, what would you like to do? Remove, Replace, or Keep". Non-words will be identified using a third party dictionary.
A third party Thesaurus or Word Frequency Dictionary with an API will be used to suggest words that have the same meaning but higher frequency in the English language. This will also work to remove idiosyncracies in the user search queries.

To address observation 2, Jellibeans will have an intermittent 'search' term box that confirms what the user wants to return. So, if the user wants to know something like lyrics of a song, or

the name of a famous person, the price of an item for sale, they would enter 'lyrics', 'name' or 'price' and this would be added to their search query (if it has not already been added by the user). This would ensure user search queries are more complete.

For observation 3, pronouns ('I' and 'You') are ignored via a stoplist which is implemented by the search engine itself. Jellibeans will also identify names of people via the third person library and suggest that these are removed, replaced with another search query term, or kept if the keyword is relevant. The extra layer of checking refines the search query.

Because Jellibeans bypasses the search engine interface, a spell checker will be implemented to address increased spelling errors in high AQ scorers (observation 5). This will also be in the form of a 'suggestion' at the search query formulation stage.

The Autism Quotient score of the user will be stored.

3.5 Implementation

What was actually done Classification of users Feature uses Storage of information

3.5.1 Integrating Jellibean with Motion Controller Hardware

LEAP motion controller and its use for the system. Current Projects Hardware Selection Process and Important Design Issues:

1. Good timing correlates to a good meaning and User Experience.

2. The leap has options to poll frames at a constant rate (to keep timing of movement accurate) which is important.

3. Cognitive lag time. Each of our senses operates with a different lag time. Hearing has the fastest sense-to-cognition/understanding time, and surprisingly sight – the slowest. If the devices interferes with the processing of the sense, it will confuse the combinatorial configuration of the senses, leading to misunderstandings in the meaning and a worse user experience.

4. Volume is important because this is a tool to be used with individuals with ASD, the device must have a low 'volume, i.e., the sensory experience cannot be overwhelming.

5. Load, by this I mean 'cognitive' load is most desirable when not high. We do not want the device to be overwhelming in terms of its cognitive load.

6. Within the selection process, I did not just consider the physical design of the device, but also the way in which the devices manifests actions into behaviours. That is, how does the user engage behaviourally within the environment using the device? What about the physical sensation and its path towards a behavioural or emotional response? For example, can we program there to be an activity followed by a reward to reinforce the activity.

3.6 Development and Implementation of Jellibean

3.7 API's and Development Tools

4 Result

The transformations that were implemented

5   Conclusion

How does it compare to the original specification This work has successfully completed aims
XXX

5.1   Signals of Quality Content

I will test and evaluate the system. Testing will involve assessing the reliability and robustness of
Jellibeans; the ease of its interaction; boundary conditions; ease of use; does it fullfil the aims of
the project. Evaluation of the system will include comparisons to existing search engines; assess-
ing how this idea can be implemented to tailor an existing systems; assessing how well the system
does compared to existing systems on a set of criteria that are only relevant to the user group in
question (a collective measure of user happiness). Evaluation will also include quantitative met-
rics such as Recall, Precision, and False Negative/Positive rates.
Apache Solr

6   Critical Review of the Leap Motion Controller

Advantages

1. Impressive

2. Uses infrared to embed the users (phantom) hand on the screen

3. New technology and novel to bring to laptops

4. East to set up

5. Has built in gestures and navigation tools

6. Can work in pretty dimly lit environments (but not all)

7. It is sophisticated (sometimes the polling frequency lets it down)

8. Picks up an impressive distance along the z axis

9. Offers a recalibration process if the controller is persistently jumpy, or there are discontinu-
   ities in the tracking data, if there are aberrations in the tracking data that occur in certain
   areas of the field of view, or poor tracking range. This can be done using the shiny surface
   such as the computer screen or mirror.

Disadvantages

1. Misses small hand movements

2. The range that it will detect is 150 degree angle along the y axis, this is reasonable but not
   always idea when gestures/hand movements are large.

3. Some parts of the screen the hands to not reach, i.e., bottom left /right of screen are some-
   times hard to reach.

4. Loses the hand, stops working/sensing the hands, even when the controller stays on.

5. Often misses frames, so the user makes larger hand movements and then overshoots (when
   the LEAP catches up)

6. Built-in controls are not ideal

7. Lighting works best when the hand is seen in silhouette fashion by the controller.

References

[1] Baron-Cohen, Wheelrigjt, Skinner, Martin, Clubley (2001). The Autism-Spectrum Quotient (AQ): evidence from Asperger Syndrome/high-functioning autism, males and females, scientists and mathematicians. Journal of Autism and Developmental Disorders, 31, 5-17.

[2] Bowler, D. M., Gaigg, S. B., Gardiner, J. M. (2008). Subjective organisation in the free recall learning of adults with Asperger's syndrome. Journal of Autism and Developmental Disorders, 38(1), pp. 104-113. doi: 10.1007/s10803-007-0366-4

[3] Developmental Disabilities Monitoring Network Surveillance (2010) *Centers for Disease Control and Prevention (CDC). Prevalence of autism spectrum disorders: Autism and Developmental Disabilities Monitoring Network, United States. MMWR Surveill Summ.2009; 58*, 10:120

[4] Novogrodsky, R. (2013). Subject-pronoun use of Children with Autism Spectrum Disorders (ASD). Clinical Linguistics and Phonetics, 27(2), 85-93.

[5] OpenSearch, http://www.opensearch.org/Specifications/OpenSearch/1.1, Retrieved 30 July 2015.

[6] Fishkin (2015) https://moz.com/beginners-guide-to-seo/how-people-interact-with-search-engines, Retrieved 3 August 2015.

[7] Surveymonkey, https://www.surveymonkey.com/home/, Retrieved 5 August 2015.

[8] Term Frequency Inverse Document Frequency Weighting, http://nlp.stanford.edu/IR-book/html/htmledition/tf-idf-weighting-1.html, Retrieved 9 August 2015.

[9] Word Frequency Information, http://www.wordfrequency.info/free.asp?s=y, Retrieved 6 August 2015.

7 Appendices

7.1 Search Query Survey

Search Behaviours

Welcome to My Survey

We are trying to understand more about how people with autism, or autistic tendencies, and people without autism search the web.

For the search queries below please indicate what search terms you would use to search the web.
There are no right or wrong answers.
Some of the searches do not have an answer but please tell us how you would go about trying to search for the answer.
If you would do many iterations of search, please indicate that below. Start with the one you are most likely to try.
Please indicate as much as you can about your search process as possible. Use as much space as you need.

There are 10 questions here and this should take approximately 20 minutes or less.
The second part asks you to complete an online questionnaire and report back the score on this sheet.

If you have any questions you can contact e.massand@bbk.ac.uk.

Next

Search Behaviours

An Example Question

EXAMPLE QUESTION:
You saw a man reading the paper today (you don't know which paper it was), and the headline was vaguely about a new discovery about the solar system. What would you type into search on your favourite search engine to find out about the story?

EXAMPLE ANSWER:
I could start by searching, 'solar system and discovery and news and today'.
If that doesn't work I could search, 'todays paper and solar system', or if I don't find my answer there I could search 'news' and manually /visually sift the page for relevant info.

Prev   Next

Search Behaviours

1.

What browser do you usually use? (Google Chrome, Internet Explorer, Safari, Firefox ...)

and

What search engine do you usually use? (Google, Bing, Yahoo ...)

[   ]

Prev   Next

Search Behaviours

2. You hear a song on the radio with the lyrics, "Look at your children", and you want to download it. What would you type into search on your favourite search engine to find out what song it was?

[   ]

3. You've lost touch with an old school friend (you went to St. Mary's School). What key words/queries would you use to find them?

[   ]

4.

How would you identify what this is using a search engine (pretend you don't know what it is called). What key words/queries would you use?
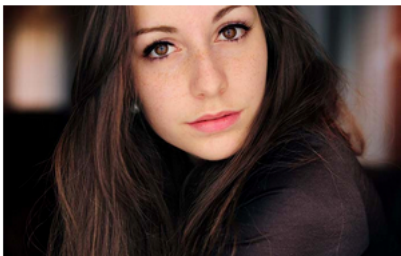
5.

How would you find out the name of this famous person using a search engine? What key words/queries would you use?

Two repeats shown

6.

What would you search for to identify this pattern's name, and which country it originates from?

7.

How would you identify what breed this dog is using a search engine? What key words/queries would you use

8. Your friend and you can't agree on how Thandie Newton pronounces her first name. How would you resolve this using a search engine?

Figure 2: Search Query Survey on Surveymonkey.com [7]

7.2 Questions on the Autism Spectrum Quotient [1]

Participants are asked to read each statement very carefully and rate how strongly they agree or disagree with the statement (Strongly Disagree, Slightly Disagree, Slightly Agree, or, Strongly Agree).

I prefer to do things with others rather than on my own.
I prefer to do things the same way over and over again.
If I try to imagine something, I find it very easy to create a picture in my mind.
I frequently get so strongly absorbed in one thing that I lose sight of other things.
I often notice small sounds when others do not.
I usually notice car number plates or similar strings of information.
Other people frequently tell me that what Ive said is impolite, even though I think it is polite.
When Im reading a story, I can easily imagine what the characters might look like.
I am fascinated by dates.
In a social group, I can easily keep track of several different peoples conversations.
I find social situations easy.
I tend to notice details that others do not.
I would rather go to a library than a party.
I find making up stories easy.
I find myself drawn more strongly to people than to things.
I tend to have very strong interests which I get upset about if I cant pursue.
I enjoy social chit-chat.
When I talk, it isnt always easy for others to get a word in edgeways.
I am fascinated by numbers.
When Im reading a story, I find it difficult to work out the characters intentions.
I dont particularly enjoy reading fiction.
I find it hard to make new friends.
I notice patterns in things all the time.
I would rather go to the theatre than a museum.
It does not upset me if my daily routine is disturbed.
I frequently find that I dont know how to keep a conversation going.
I find it easy to read between the lines when someone is talking to me.
I usually concentrate more on the whole picture, rather than the small details.
I am not very good at remembering phone numbers.
I dont usually notice small changes in a situation, or a persons appearance.
I know how to tell if someone listening to me is getting bored.
I find it easy to do more than one thing at once.
When I talk on the phone, Im not sure when its my turn to speak.
I enjoy doing things spontaneously.
I am often the last to understand the point of a joke.
I find it easy to work out what someone is thinking or feeling just by looking at their face.
If there is an interruption, I can switch back to what I was doing very quickly.
I am good at social chit-chat.
People often tell me that I keep going on and on about the same thing.
When I was young, I used to enjoy playing games involving pretending with other children.
I like to collect information about categories of things (e.g. types of car, types of bird, types of train, types of plant, etc.).
I find it difficult to imagine what it would be like to be someone else.
I like to plan any activities I participate in carefully.
I enjoy social occasions.

I find it difficult to work out peoples intentions.

New situations make me anxious.

I enjoy meeting new people.

I am a good diplomat.

I am not very good at remembering peoples date of birth.

I find it very easy to play games with children that involve pretending.