26-11-2021

~Teun Visser; 442651

Fontys; university of applied siences

What is a reliable way to identify users, without their direct input

Pollstar

Contents

[Main Research Question; 2](#_Toc88830617)

[Sub questions; 2](#_Toc88830618)

[Research Method; 3](#_Toc88830619)

[Available product analysis 3](#_Toc88830620)

[Community Research 3](#_Toc88830621)

[Expert Interview 3](#_Toc88830622)

[Sources; 4](#_Toc88830623)

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Changes;** | **Author** | **Date** |
| 0.1 | Setup, Creation first Questions and listing methods | T. Visser | 11/25/2021 |
| 0.2 | Added Brainstorming part, in *Expert Interview* research Method | T. Visser | 11/26/2021 |
| 0.3 | Community Research, Available products researched | T. Visser | 12/20/2021 |

# Main Research Question;

What is a suitable way to identify users in this project scope, without their direct input?

## Sub questions;

* What data can you get from website visitors, by just loading the site on their preferred browser?

I won’t be query users for information, so I have try to get as most from the browser supplies me with, like; IP address, I could save something like a cookie, etc…

* What about privacy concerns and cookie notifications and performance?

As mentioned above, I plan to collect data, but I’ll have to look into how much I can legally collect without asking verification from the user or slowing the application.

# Research Method;

## Expert Interview

As mentioned in the other methods, this problem has several solutions, varying from saving and retrieving small pieces and guessing, our entire algorithms that creates and compares the online fingerprint of the users. It is very hard to know where to start searching and where to stop, so I’ll be asking some questions to the one who grades and judges my final implementation. The things that I want to know is; how and where to start? What are your thought and experiences? And where to end, what would suffice? Using these answers, the research and implementation following that would be a lot less rocky.

## Community Research

After checking quickly online, I am not the first one with this problem and others have tried different solutions, but every solution has its pros and cons. I shouldn’t try to reinvent the wheel here but try to learn and understand if other people their solutions are just applicable for me.

The results here will be varying, because answers change based on the question asked, I’m not trying to bite of more than I can chew. I’ll be looking at solutions that are effective and good for a prototype and it should be able to integrate with the plans for the backend.

## Available product analysis

This project idea isn’t an original one, and it has been done on several ways, all with difference in usability and functions. They all share the same core with the same idea, so that means my specific problem here should have been tackled already, or at least a version of this.

The plan here is to try these similar services and see how they handled it, from a user’s perspective, at the same time I will be trying to break their solution and find a way around it.

This method won’t only give me a good idea of how the end project could work, but also provides me with a steady baseline of the competition, how far should I go with this implementation until it costs more then the payoff is worth?

## Expert Interview;

### Brainstorming

|  |  |
| --- | --- |
| Email verification | |
| Pros:   * Reliable * Fool proof | Cons:   * Can be automated with bots * Doesn’t comply with the Quick & Easy user Story |
| Cookie (saving uid on client side) | |
| Pros:   * Complies with the Quick & Easy User Story * Easy to implement and to use * When voter upgrades to signed up User, their previously voted on polls convert with them | Cons   * Little technical skills needed to work around and break * Users will accidentally find work arounds without extra input |
| IP address saving | |
| Pros:   * Complies with the Quick & Easy User Story * Makes sure users can vote only once, even with different devices (on the same ip, of course) | Cons:   * Several users on 1 IP will give problems * Can be worked around by using vpn |
| RCN Fingerprint | |
| Pros:   * Complies with the Quick & Easy User Story * Very hard to Crack & Hack * Will definitely makes sure I graduate | Cons:   * Incredibly hard * No working demo’s for my Use Case, so hard to know if it works beforehand |

### Conclusion

A more grandiose method would of course be amazing, but due to time constraints, limited resource and way to much other stuff to do, I won’t be looking into methods of this magnitude.

For the Pollstar project, the implementation of this research question with it’s accompanying user story should be kept at a beginner level, it should work with a specific set of front end tests at the end of creating the demo, but it doesn’t have to do more then prove the point of how it would work in an ideal situation.

## Community Research

|  |
| --- |
| Input: “how to differentiate users website without login” |
| <https://www.reddit.com/r/webdev/comments/6qn5b4/identify_users_without_login/> |
| pilibitti:  *What you are asking is sort of the "holy grail" - frictionless login and it is kind of impossible.*  *You can use cookies of course, act like the user registered and open a session for them. If they choose to do so, they might create a real user on the site then you link the temporary persona and the new user and all becomes well.* |
| wordaligned:  *Maybe a captcha and a session cookie will provide the balance you're after? Difficult to get past the first hurdle for bots, convenient for subsequent votes on other images.*  dweezil22:  *OP's specific requirement is just to mitigate spam votes, and a captcha will do that (and just that).*  *If OP isn't willing to accaept the case where a human being manually votes multiple times, then a real auth mechanism is required, with the user alienation that will accompany it.* |
| Summary:  OP has a situation similar to mine;  suggested solution is to pretend that website visitors are actual users, and create a cookie. when said user wants to signup the cookie is assigned to the user.  This will make it so that users without malicious intent can use the site like intended.  to counter the use of bots, another user proposed a cookie in combination with a captcha, OP found out that there are invisible captcha’s available from Google’s api.  If the captcha could not be solved arrhythmically it would need user input, but it might be something to consider. |

|  |
| --- |
| Input: “How to uniquely identify a web user” |
| <https://devrant.com/rants/201433/i-need-to-figure-out-a-way-uniquely-identify-each-computer-which-visits-the-web> |
| BoyBiscuit:  *This should help you with JavaScript identification https://github.com/Valve/... it's a rather heavy solution compared to some other techniques but it has a high accuracy* |
| Summary;  OP here wants specifically no cookies, a user here suggests the use of fingerprint authentication, on the official site it states that the identification rate is quite high and rather reliable, it’s harder to spoof than a cookie but should still be within the ease-of-use requirement. |

|  |
| --- |
| Input: “Cookie notification requirements” |
| <https://www.termsfeed.com/blog/cookies-notification-messages/#:~:text=A%20cookies%20notification%20message%20is,land%20on%20your%20homepage%20first>. |
| There are 2 main laws that affect cookies:  EU Cookies Directive:   * Owned by EU businesses, or * Directed towards EU citizens   The main requirements under this directive are that:   * Users are informed about your cookies usage, and * You get consent to place cookies before doing so   General Data Protection Regulation (GDPR) applies to Websites that:   * Offer products and services to citizens of the EU, or * Collect personal information from citizens of the EU   GDPR requires that:   * Websites get active consent to place cookies * Users are able to easily withdraw consent and opt-put of cookies placement |
| Summary:  There are different types of cookies; and they have different rules:  The GDPR applies to cookies that are used for advertisement, analytics and other services that collect personal information; in this case you need active consent.  For smaller use-case cookies only the EU laws apply.  the difference between these to is that the GDPR rules require to actively consent to the type of tracking cookies, while the EU laws only needs users to consent to the use of cookies in general. |

## Availableproductresearch

What already exists;

<https://strawpoll.com/>

Vpn ips are filtered out and are not allowed to vote

on about page; “We already have bots and scripts under control”

<https://www.poll-maker.com/>

We employ several evolving digital fingerprint technologies which monitor voter trends. In addition to this unusual responses on large polls are flagged and reviewed manually. To this end we collect IP address, browser, referrer and other details which help us to evaluate the authenticity of a response and evaluate trends.

<https://www.strawpoll.me/>

Has several selectable options;

* + Selecting Normal Duplication Checking is the most secure method to ensure vote accuracy. It will block all duplicate votes based on the IP address of the user.
  + Selecting Permissive Duplication Checking will block duplicate votes based upon the browser of the user. This does allow multiple votes from the same IP address. However, the users would have to vote from separate devices.
  + Selecting No Duplication Checking is the least secure method to ensure voting accuracy. It will disable the duplication system all together. This means that users can vote as many times as they would like.

<https://doodle.com/poll-maker>

has no information on the site about anti-fraud measures, but it appears as they are tightly integrated with services like Microsoft and google, requiring you to login to vote. This is effective as anti fraud and the pollster can easily detect if there are emails that were not intended to vote.

<https://directpoll.com/>

Directpoll has taken an all or nothing approach;

You can set the poll to open, this will make it available to everyone, but it doesn’t appear to have any special measurements to detect malicious votes.

On the other hand, you can create a restricted poll, with this you can create an amount of tokens, that is the maximum amount of votes said poll can receive. These tokens can be distributed so you can be sure that only the people you give a token can vote, once per token.

Summary

Every site takes different measures, that gives them all a different appeal. Some of their solutions I will be taking inspiration from, but also some of them will be out of the scope of this project; due too time constraints and/or the fact that they directly clash with some of the requirements.

## Results

Expert interview;

In the expert interview, several methods rose up that could be used, some of those simply don’t fit the scope of the project.

The biggest thing to take from the advice and input from mr. Reemer is that there are a lot of ways of accomplishing this goal; every solution has reasons why it would and wouldn’t be a good contender for the main question it’s answer.

This doesn’t mean there aren’t any solutions with drawbacks that fall within the scope, it comes down to comparing pro’s and con’s while not only keeping the project’s requirements in mind, but also the technical knowledge kept in mind.

Community Research:

The community had varying solutions to the stated main question, one of the solutions is to use the online fingerprint, this kind of solution claims a high success rate. There are multiple solutions that work on the same concept but use different input for different use-cases, but one of their biggest cons is they are quite a grandiose solution.

There are also a lot of solutions that are easier to grasp the concept of, although they have varying results. The one I investigated and listed in the research matches the main question, it will differentiate users and protects against bots.

Cookies might be a problem; Using cookies without the user’s permission is going against some laws, and to get permission users have to explicitly give permission to the use of cookies, this will be an extra button.

Available product Research:

The comparable products can be splitted up;

Quick methods and thorough methods

Quick methods ask for cookies permissions and immediately allow poll creation and voting.

This allows for quick action; the filter works by using fingerprints or IP validation, which uses no input.

Thorough methods have bigger drawbacks; some require a sign up or sign in with 3rd party options (like google or Microsoft) to create a poll and vote on one.

Sub questions;

* What data can you get from website visitors, by just loading the site on their preferred browser?

In my research I found a lot of items, such as the width of their screen, operating system and language settings. These items on their own are too general to pin point a user, but combining all these facts creates a digital fingerprint from the user, this is hard to imitate but at the same time hard to get rid of your own.

There is also your IP address, this can be included in the E-fingerprint, but it’s more reliable, they are unique to a group of users and can be detected when spoofed

There is one more thing, a website can save items to a device without the need for a user input (except for permission). It’s not too hard too work around and cheat, but for users without malicious intent it’s perfect.

* What about privacy concerns and cookie notifications and performance?

When a site wants to retrieve or save items to a target device, they must obtain permission from the user according to the EU Cookies Directive and GDPR laws. Users must explicitly accept the fact that the site wants to store and retrieve data from the user.

This clashes heavily with the user story from the project that states: “When a poll is opened, you can immediately vote”.

This means that I cannot use cookies that fall within the GDPR laws, as that would require a dedicated button to consent. But if I stay away from those laws and the cookies that surround those, such as personalized ads cookies or information gathering ones, only the EU cookies law is triggered. The EU law still requires consent but the requirement is easier met, it doesn’t directly need extra input.

# Main Question

What is a suitable way to identify users in this project scope, without their direct input?

Cookies.

The provided research came up with a lot of different answers, but this one fit requirements for the project the best.

Cookies plus the google reCAPTCHA will comply perfectly within the scope and fit the bill of \*no extra input required\*

For users without malicious intent, they can vote without hindrance, but under the hood is confirmed if he is an actual human, and his cookies make sure this user can only vote once.

If the user has malicious intent a cookie can be easily evaded, but cheating votes will be hard to automate, because of an invisible reCAPTCHA. A user could of course manually cheat votes, but the project requirements state that it’s intended as a quick polling service, if users manually add vote it won’t make a enough of a dent in the final results.

Other noteworthy solutions:

IP address: this is unique per router; that would mean that only 1 person per organization is able too vote. Although this is a tradeoff and would be really useful and probably worth it, an IP address is considered personal data, and the collection of this would need the user to consent to GDPR laws of cookies… which requires extra input.

RCNA Fingerprinting; this would be perfect if it wasn’t so hard, the concept is easy to grasp, but only integrating enough for a prototype polling site would make it unreliable, full building of a fingerprint is out of scope for a prototype of a year 2 HBO student…

The Research itself

The found information and resulting conclusions fit the requirements well, mostly all the results are relevant to the main question and or sub questions. The recorded information is supporting the research and is relevant to the sub- and main question. But it’s lacking. There is much more information to be gathered from competitors and the community was brimming with alternative ways of handling the given problems. But a lot of results are excluded because of assumptions made about it’s relevancy to the main question and the project scope, and due to time constraints.

Some suggestions for future work on this topic

Use cookies as a baseline, but add another solution to this; so using cookies in combination with an IP; you could still build a profile for every user, and include the IP in there. The results of the poll then include where the vote came from and most importantly can help detect fraudulent votes by identifying duplicate votes or VPNs.

An even better solution would be to implement several solutions and let the poll creator pick what fraud detection to use, as found in the Available product research.

# Sources;

<To be Formatted>

Research Methods;

<https://ictresearchmethods.nl/Methods>

Frenk Reemer, Teacher at Fontys, University of applied siences