Wikifunctions feedback

This research report is based on interviews with 10 programmers. Only three were professional developers. The rest used code in their job or at school. Participants represented the fields of biomedical engineering, bioinformatics, geophysics, mathematics, statistics and actuarial science. They reviewed updated examples of the Wikifunctions UI.

The goal was to interview people who already knew how to code but did not think of themselves as programmers, at least not professionally. This could bring a diversity of perspectives to the project and reach beyond the developer subculture.

Project Strategy

- ➔ What does the user do with these functions?
- ➔ How do we evangelize functions?

System Architecture

- ➔ Maybe functions are used like templates
- ➔ Are individual implementations addressable? How do you select among multiple implementations when executing a function?
- ➔ How do you run/test an implementation that doesn't run in a browser? FORTRAN or C++ emulators? MATLAB or Stata licensing?
- ➔ Graphing / visualization functions
- ➔ Github vs MediaWiki
    - Github has a learning curve
    - Github is for complex software; not simple functions
    - MediaWiki version control

User Interface

- ➔ Does the interface scale for complex functions?
- ➔ Are complex functions out of scope?
- ➔ What if there are hundreds of implementations for a function? With multiple implementations per language?
- ➔ How to sort/filter/triage all the functions? Which function is best? Which implementation of the function should I use?
- ➔ How to edit the source code for a function?
- ➔ How to create a new implementation?

Project Strategy

What does the user do with these functions?

Ian: I understand all about the objects and the function structure, and I know what a lambda is and all that sort of thing. **But how do you use it all?** Because I've been through documentation to some extent, and it's slightly confusing. I have been unable to find a conceptual high level view that explains why you're doing it.

Ian: Having all these functions, how do you actually use them as an end-user, or as a Wiki page editor, **how do I invoke** all of these things?

Ian: Normally you'd start with a user need, but I'm seeing Wikifunctions actually starting the other way around and saying, "Okay, well we can make functions," but **how do you make use of them?** Do you see my problem in effect there?

Stelio: **You should start with the use case**, and then from that, build up what the interface needs to look like, as opposed to thinking, "Well, here's an interface. Can we shoehorn every kind of program into it?"

How do we evangelize functions?

Andrew: But I am wondering, what would it take for somebody to be entering content within a Wikipedia article rather than just an info box? W**ould anybody actually want to do this kind of thing**, have a function that's giving you the information, rather than just type it straight in?

Andrew: The question would really be **how many people would even know** enough to realize that they should look for a function?

System Architecture

Maybe functions are used like templates

Ian: When I first heard about WikiLambda I thought, "okay, this is fine." I get the point. But **what I want is some worked examples** that go all the way back to the user level and say, okay, what actually happens?

Ian: It's a Wikipedia page, and somewhere in that I'm calling functions. How do they surface, right? **Is it in effect like, a template would be in Wikitext** instead of being implemented in the templating language and possibly Lua, am I going further back and having the language at the back calculate something?

Andrew: The other thing that I wonder about is **how this would be used within Wikipedia**? Would it be one of these things where you're converting units? We've got various templates for doing stuff. Is it that sort of thing, that you'd basically be maybe in the long run trying to put that into articles?

Tiago: How would this be **embedded inside other Wiki projects**? Like, how would they use Wikifunctions in Wikipedia? Would that be possible?

Are individual implementations addressable? How do you select among multiple implementations when executing a function?

Ian: If you've got functions implemented in more than one language, **how do we choose between the languages**? I mean, run the function, in some sense? How do I select which implementation I'm trying at this point?

How do you run/test an implementation that doesn't run in a browser? FORTRAN or C++ emulators? MATLAB or Stata licensing?

Stelio: In terms of testing it and putting values in and getting out, obviously a lot of these are super quick, no problems with them. **You can literally code them in a MediaWiki page** anyway, just using the parser functions that are built in.

Stelio: But some of the potential things that are worth building are expensive in terms of runtime, and that's going to be a question that needs addressing. Do you just give the example of the code, or **do you also provide the interface to run the slow code**? And it might be enough just to say, "Here is the code. Take it and run it on your computer," and then people can try that, for really slow functions.

Stelio: **We can't limit ourselves to just a traditional programming language**, because the definition of programming language is imprecise, and I think Excel is a very useful thing to have on here. But how do you display that within the browser? It needs, essentially, a new type of interface.

Joseph: A lot of what you would write in MATLAB are actually just functions and scripting. In terms of running the script I think **you'd probably have to have some form of emulator**. You need to **purchase a license** for the service to be able to run it on your backend.

Ian: **Is the function being actually calculated at the server end**, **or is it being calculated in the browser?** And I guess in theory, it could work with both. So you'd need a server side language and a browser language. Now, JavaScript could be both of those. Or have PHP at my end and JavaScript at the other end.

Darran: I would presume the same for Wikifunctions, that if I had to do the testing of the code, **there would be a way to do it onsite or in the site**. You can go in and test yourself.

Graphing / visualization functions

Joseph: There's a lot of **graphical computing** that you can do in MATLAB. It definitely would be useful in some cases. But it really depends on what kind of support you have on the backend.

Joseph: You can definitely do it with high level languages like Python, with built in libraries. **I guess it all depends on how you support it.**

Stelio: Image-based functions would **absolutely have a place here**. A very simple example would be, given an array of numbers, output that array as a bitmap image, where each cell number represents a color pixel. That's a straightforward and elemental function that I think would absolutely sit here.

Stelio: The interesting thing would be, how do you serve that content? How do you set up the interface to put the inputs in, and how do you set up the interface to get the outputs of it out? You might have something where you want to pass in an image or a sound file or a video, and these become progressively bigger and bigger objects, and **you don't necessarily want the Wikifunctions server to have to cope with the strain of doing video editing** in real time.

Abraham: If this site were to offer **even a basic visualization**, a curve for example, that would be great. For most simple data, it's just two axes, X, Y. That does the job for the majority.

Abraham: It would be great if you could give people to **download the data** in a format that they can use other visualization software.

Github vs MediaWiki

Stelio: **GitHub would be great for the higher-level programmers**, and I think MediaWiki would be better for the lower-level programmers.

Stelio: In a sense, though, I would have thought that the two could communicate with each other. It wouldn't be impossible to have a **MediaWiki front end with a GitHub back end**, and for the GitHub back end to be accessible for the high-end user who wants to go in there and go directly to the GitHub back end.

Stelio: Having two layers, I think, has merit. MediaWiki, to be able to easily find what you want to, and then to have a link to GitHub directly if you want to then go that route, and that would be the **best of both worlds.**

Darran: **Combining Github and MediaWiki** sounds good.

Github has a learning curve

Ian: I use GitHub daily through my job, so I'm familiar with the thing. But I suspect a lot of the people coming to Wikifunctions are not necessarily programmers. In order to work with GitHub, **you need a whole new set of metaphors and ideas and concepts that aren't really present** in the Wikimedia environment. Adding GitHub on top of it would be just another confusion.

Darran: I dabbled in GitHub myself. So anybody I think who is actively programming has to know their way around GitHub just to build up their own portfolio. For a casual programmer, **I don't think it is such a huge learning curve.**

Tiago: I'd say that perhaps half of the computational biologists don't use any kind of version control, which is amateurish. **Github would be a limitation**.

Charles: It's getting more common. But it's still the **distinct minority** of scientists or mathematicians using GitHub.

Github is for complex software; not simple functions

Ian: Only the **biggest and heaviest of the scripts** go out to GitHub.

Joseph: This is really just providing functions in various programming languages. But i**f I'm looking for a software package, I'd probably be looking at GitHub**, but if I were looking for a function, I'd probably use something like this.

Joseph: Functions are standalone, but **anything that's more complex that has dependencies is probably going to be on GitHub.**

Joseph: There's tons of bioinformatics software on GitHub. When people are using bioinformatics pipelines and whatnot, usually it's going to be just **calling it straight from a GitHub repository** if it's just plug and play.

MediaWiki version control

Ian: Version control in MediaWiki is absolutely fine. **It would be more of a trouble to use two separate systems.** You've got everything in Wikipedia, then you've got the history of the page without any real problems.

Darran: With GitHub, **it's built for that purpose**. But to put the functions on a Wiki. And it's a little bit kind of squeezing it in for the sake of this.

Charles: You can have **edits get lost in the Wiki model**.

Charles: It's very easy to accept useful, minor edits, "Oh, I fixed the typo for you," and major useful edits, "hey, I added a new feature for you." And it's easy to protect major bad edits. If someone goes in and scrambles things around a little bit, but not enough that something really overtly breaks, **what's the chance that it gets fixed**? Maybe it's 80%, but definitely not 99%, or 100%.

Charles: With Wikipedia, that's largely balanced out by the fact that it doesn't matter so much, and not that many people are malicious. But, I wonder with code. I'm not saying that's a show-stopper, but it's a **reason to be concerned**.

Mahir: With Github it's possible to store individual granular changes and then reapply them. **Version control for software is amazing**, and I'm hoping that if in the interest of taking advantage of the efficiencies that that imparts, that Wikifunctions uses them in a lot of ways.

User Interface

Does the interface scale for complex functions?

Ian: Obviously, you start getting bigger and messier functions, **you're going to have real estate issues on the screen. And your nesting's going to be kind of deep and horrible**. So you probably want to mock up something really dirty and large to see how it would end up scaling.

Stelio: It might work fine for simple examples like multiplication. **I'm not sure how you would do an interface in this style for something more complex**, like, say, a sorting algorithm. That's a harder thing to show. And therefore, I wasn't sure how useful this kind of interface was going to be, unless this is just defining the inputs and the outputs of the functions.

Stelio: I think this interface works for **very, very small stuff, but it won't work for anything that will be a real thing that people want to look up**, because typically people won't want to look up, how do we take two numbers, do something with them to get one number. That's so basic that they probably already know how to do that.

Stelio: The problem will be a real-world problem, and **real-world problems are messier than the very small examples there.** It's not, "How do I go from an atom to a molecule?" It's, "I've got a crystal structure now, and the structure of the molecules within that, there's a little bit of chaos in there."

Joseph: In terms of bioinformatics, probably really basic algorithms that you can just run as a function, you can probably put on this. **Anything more complex, I don't think a platform like this would really support well.**

Are complex functions out of scope?

Stelio: This should have all the different functions you could put together to make a program, but it's **Wikifunctions, not Wikiprogram**. It's not, "Here's a complete solution, here's a complete package to your problem." It's more, "Here are the tools you need."

Ian: Conceptually, **I'd like to think that all functions are fairly small** and that they're calling other functions, but you may find that you've got performance issues if you are nesting that big. So maybe for purely pragmatic reasons, you want to put more of the code in one function.

Tiago: **There are some simple functions that would make sense** to be written in Wikifunctions, but I am not sure.

Andrew: The vast majority of scientific knowledge, it's either a straight number or it's an equation that could have anything in it. **That probably is way too complex, especially for the first crack at this.**

What if there are hundreds of implementations for a function? With multiple implementations per language?

Stelio: Obviously there are **multiple implementations of a particular use case**, because you have multiple different scripts. Over time, you'll get enthusiasts adding in their favorite script for all kinds of things, and you'll end up with a list of hundreds of different programming languages available for each one.

Stelio: There will be **multiple versions for each language** as well, because one language will have more than one way to do a particular thing which varies by use case. What is the quickest code? What is the shortest?

Stelio: It's going to be a question for moderation, in terms of, how do we cope with it when multiple users want to add in their version of a function? Whose do we choose as being the best? **Custodianship of the site** is going to be an interesting logistical question.

Stelio: There are good reasons to have multiple implementations in some circumstances. But then you have the issue of that if one user sees, "Oh, well, there's three implementations of this, but none of them are like my favorite, I'm going to add my favorite too," is their favorite worth adding? **Does it add value to the site?** And that's not a trivial question to answer.

How to sort/filter/triage all the functions? Which function is best? Which implementation of the function should I use?

Stelio: I could easily imagine the list of implementations being huge, and we'd pr obably want **some way to collapse that down, or filter it**. e.g. 99-bottles-of-beer.net

Stelio: If I come here as a user and I say, "Well, I need to sort some data," and I come onto Wikifunctions and look at the category of sorting functions and I go, "God, that's loads of them. **Which should I use**?" I'd want to be able to have the information that would help me, as a user, decide that.

Stelio: There's got to be at least a **dozen different sorting algorithms**. I would want to know the use case. Why would I want to use this sort algorithm as opposed to that sort algorithm? This one works better with text than numbers, for example, or if you've got partially sorted data.

Mahir: We might want to have a way to **identify limitations within any particular implementation**. Like maybe there might be an icon next to Javascript saying, this does not support big numbers or this does not support, I don't know, complex conversions

Mahir: It would be useful to see at a glance **what does an implementation do that another doesn't** so that we can suggest that to a potential coder. Seeing what can be improved as far as the functions themselves go.

Mahir: **Having a way to rank them would be important**. Because presumably, if there are two implementations built on the same language, there has to be something in mind as far as maybe the trade offs that one has over another for that same language. Which is the one that we should look to first?

Joseph: The categorization that you have on the main page is perfect. It lets you narrow it down by topic, but if people are searching for a particular topic… they know what kind of function they have in mind, **how would they find it?**

How to edit the source code for a function?

Tiago: **What's weird there is that I don't see the function itself.** Maybe not the code exactly, but some indication of the function.

Darran: I'd like to get more information. **Where's the formula?** Maybe, the Javascript and the Python and so on can be expanded. And I don't know what happens when you click on one of them.

Mahir: Sometimes we might need a more complex set up before trying to test a particular function. There should be an option to sort of go in and **specify with more fine granularity** what kinds of things need to be prepared in advance.

Stelio: The algorithm in the middle, I don't know that the algorithm could be defined in such a simple interface, because the algorithm ultimately is going to be **the block of code itself** in the different programming languages.

How to create a new implementation?

Darran: Where can you **do this in a different language**? It seems very straightforward for people to put functions together, but if you wanted to go in and create a similar function in JavaScript or another language, it'd be interesting to see how to do that.