Crowdsourced recognition of recoil black holes

Bachelorarbeit im Fach Informatik

vorgelegt

von

David Arpad Pinezich

Zollikofen, Bern, Schweiz

Matrikelnummer 11-729-464

Angefertigt am

Institut für Informatik

der Universität Zürich

Prof. Abraham Bernstein, Ph.D.

Betreuer: Patrick de Boer

Abgabe der Arbeit: 14.10.2015

Crowdsourced recognition of recoil black holes
Bachelor Thesis

David Pinezich
Department of Informatics
University of Zurich
Binzmühlestrasse 14, 8050 Zürich
Switzerland
david.pinezich@uzh.ch

ABSTRACT
This paper is focused on the "Blackhole Chaser" a crowdsourcing platform prototype which helps to find new recoil black holes. First this paper shows a general overview over the topic of crowdsourcing and other related research fields. Then it explains how the "Blackhole Chaser" was built and how the architecture was planned. After that, it describes how different users (paid crowdworkers, it-specialists and professionals) act on this platform and if there is a coincidence in their classification on it, based on their individual cognitive skills, that have been tested prior to using the platform with the renowned ETS testing framework. This paper closes with a conclusion and a listing of future tasks.

General Terms
Algorithms, Measurement, Design, Human Factors, Verification.

Keywords
Crowdsourcing, Crowdworker, Global Brain, Recoil black holes, SDSS1131, upWork freelancers

1. INTRODUCTION
If we have a look at todays computers, it is usual for us to have them in a network. We are well aware that not all knowledge is saved or calculated by the computer we have at our own place. It is for us a fact, that some of the big calculations happen on bigger computers than ours or even on multiple computers or servers.

But still we know about tasks which are thinkable not easy to solve for a computer even in this size. That is because it has no database for specifying something as “human natural”. That is something humans can still do way better than computers, making decision which are not only statistically true, but also have to “feel right” [1]. This paper will first show with a literature review how this “feel right” with crowdsourcing and near researches was brought to a whole new level. How the global brain [26], which was first an idea, became with micro task management suddenly reality, and what are the weak points of crowd-sourced tasks or crowdsourcing in general and finally how to overcome them. With this knowledge there will also be built a link to the actual SDSS1131 inspired research. It should also be noted that this platform is inspired by the well-known Galaxy-Zoo platform. After that, there is a detailed explanation of the idea of this platform and how it was actually built and tested with users. It will deliver a deep look into the frontend which is the most important part so far in this web platform. According to that, the use of this web platform by actual users will be discussed and with three research questions about the platform it will be shown if the platform is done for everyone or if special cognitive skills are needed / or help during classifications. For the next part, all gathered research results will be combined and discussed. And for the last part, it will be shown how the future work on this platform looks like and which ideas and tasks are still open after this study.

2. Literature Review
When it comes to more difficult tasks like the classification of images for special abilities or the decision of how “good” music is, is the human brain is still superior to the computer. For us humans this is clearly an easy task because we have always the feeling if something is right or at least if it seems to be “going right”. We do not have to think long about something and make statistics about it because our internal decision systems connect knowledge and already experienced situations in seconds to find a way how to solve our next task most suitable [1] in a natural way. If none of these fits, we still have a feeling for the best thing to do, while a machine is only able to develop heuristics and learn from the past or approximated future. We then moved further onto how to generalize this processes which are so natural to us and came to the famous process handbook from Malone et. al. which delivers us recipes about how to structure a process which is more in a human kind and later on the same is even doable by a machine since the process is so well structured [2] or in other researches how to do it by “knowledge clustering” and therefore with an intelligent process selector [4]. Malone et. al. then continued in their research and started to study the coordination of such processes and how to optimize them. Nevertheless, it was only 1994 Malone had already a model which tells us that such a standardized process is best to be achieve in a crowd-like surrounding [3] where over- and underachievers can equalize the general output. Some years later Bernstein et. al. described how to fulfil a process when something is problematic or not working anymore with cooperative working tools [5]. Even then again we faced that the human factor is still important and our skills will never be 100% replaced by a machine but we can already see that working together is giving us a much better effort than working alone and in single tasks. So why should we not combine the “natural thinking” of a human with the efficiency of a power-computer?

It took then some years until the market of computers and network went so cheap that almost everyone had an own computer at home and internet available. With this setting the idea of the “Crowd” and even Crowdsourcing was birth again. It was now easy to connect a lot of people at almost the same time to work on the very same certain task. Bernstein et. al. then researched this thematic deeper with their Collabio game were you can tag your friends and then see who is the best networked person in your circle of friends [6]. Sheng et. al. then found with their “get another label?” [7] that a repetitive process improves a certain task for a group dramatically. In other words, doing the same thing over and over again improves the general result. Which brings us close to the state of the art crowdsourcing. Malone et. al. then developed this idea further and showed how Create&Decide can help us in such crowd sourced

questions [8]. With this new influences Bernstein et. al. developed their project "Soylent" which is a word processor which is driven by a crowd inside [9] and showed us that even if we have a large group of people with some members which are very lazy and some which are "eager beavers" we still will have overall a good result because the crowd stabilizes itself.

While crowdsourcing and over all crowdworking gained more and more attention even more data was created at the same time. With such a huge amount of data, another important research path was born, the one of Data-Mining. Kietz. et. al. then soon developed a workflow how to mine the data with an intelligent workflow [10] and Serban et. al. even did an automated experimentation out of this and searches for the best workflow for a specific Data-Mining problem [11]. With this, more and more data and also more and more network based tasks are established Bernstein et. al. asked then themselves in a two-week experiment with 60 candidates how to develop a FeedReader add-on with the focus to tell your friends about feeds they might like. In this experiment the recommendations were important and also the social awareness and feedback of the users. They developed that people who use feeds like to get to know about new feeds but the "sharer" does not want to annoy his friends with useless links [12].

As it is already discussed, large groups of people can do tasks which are easy for humans but not so easy for computers very well if the process is extracted and the steps are repetitive and self-controlling. With the new technological possibilities, it was only a matter of time until the "Citizen Science" was born. Instead of having a researcher now occupied for a year with "viewing" of data which is important for his research. He now rather extracts the tasks which is not doable for a computer and establishes it as a crowd task. This is at the same time a huge win for the researcher and fun for the user which play its part. On the well-known website of Galaxy-Zoo (http://www.galaxyzoo.org/) exactly this paradigm is used.

We have here a huge amount of images of the universe which can have special occurrences in them, which are worth looking deeper into it. But it would need a large amount of researcher to get through all this images to have them viewed. That is when the citizen scientist comes to play. Since it is easy to extract the needed data and Galaxy-Zoo managed it very well with their User-Interface (UI) to show what we they are looking for, even non-scientists were now able to view the image and tell if there is such an occurrence of a special pattern or not. Since citizen scientists are not mandatory professionals, rather interested people we have the problem that the amount of false classifications is significantly higher than it would be if they had been made by a professional.

But this is were the crowd is superior in regulating itself. With having this images of certain parts of the universe not only shown to one or two citizen scientists rather shown to hundreds or thousands, the error gets really small and vanishes almost. And while this task is repetitive and kind of boring for a professional people outside of this research, the described "citizen scientists" love to take part in this topic because for them it is something really new and exciting. It is then no wonder that the interview with Galaxy-Zoo users of Raddick et. all. was a huge success [13] and users told them that they love to take part in research of some "new" things which they partially not even understand.

While people are eager to work on a project when it is new or something special as it is in the Galaxy-Zoo example, there is also another important fact which motivates people to do such tasks since all time, money. With Web platforms like Amazon MTurk, eLance or the newer upWork, we have now a certain meeting place were it is only mandatory for a user to have a steady internet connection to solve tasks and earn money.

Little et. al. developed the MTurk concept further so that you can easily do a process with their TurKit and then let it be done by a defined amount of people eager to get the money for an easy task which is may or may not really boring and repetitive [14]. A little bit earlier Franklin et. al. established their CrowdDB which combines a regular database with crowd elements. So that some of the tasks are probably done directly by the database engine, and some are done by a connected crowd automatically [15].

With the concept of earning money without detailed knowledge and also the experience of something new and not yet researched we can see that there is a huge amount of people willingly to do such repetitive tasks for research and money purposes. But the crowd which is established here does not have any kind of rules of their behavior and there for needs again a lot of guidance by a professional. If we switch over to the example of Jabberwocky Programming Environment of Ahmad et. al. we could get the impression that if we split the task to a lot of microtasks there should be almost no cost on managing the crowd [16]. But as we can clearly see, the crowd only works if it is well selected and the workers have some kind of Judges which guide the workers into the right direction.

As an example for this we take Climate CoLab where it is the purpose to send-in proposals or ideas to solve a certain problem which occurs at the moment. As an example: a certain country has a problem with the quality of their water, the crowd then develops ideas to solve this problem (theoretically) and has to publish their ideas for the judges individually. Since there is also a decent amount to earn if your suggestion gets picked, there are also a lot of people sending in suggestions which are not likely to have an effect at all, but maybe have a chance to get a slice of the money.

With this we can see more clear into the crowdsourcing topic, examples have shown even if the crowd is a huge self-regulating group of people which is best managed by microtasks it is really important to have on one hand people which can manage the progress of the crowd or solve their problems and on the other hand the description of the tasks and the UI has to be so easy that literally everyone can work with it. In the example of ClimateColab there were metrics established to see if a suggestion is good or rather not so good and also the human factor is taken in place with people pre-judge suggestions before they are seen by the actual master-judges [17]. Zhang et. al. also tackled this problem and tried to solve it with global constraints. And here again we see that a big crowd solves the error of a single person itself and also that a strict and easy to understand how-to for solving the task is always needed [18].

With this knowledge about the crowd it is well worth noticing the research of Weld et. al. which established an online education crowdsourcing challenge where one-to-many remains still the loved art of teaching but with the here established crowd which is aimed to review the work of the student, the feedback for the student gets broader and content-richer for the student because sometimes one can read the same text written by two individuals and just understand one of them way better [19][25]. An other view is the work of Minder et. al which established the CrowdLang, a programming framework which is able to combine all this processes and control mechanisms and then let the task be solved by a self-regulating group of MTurk workers. After a broader evaluation of them they found out that a Double Six Sigma Pruning has scored the best results used with their CrowdLang [20] which

now affects their ongoing studies. One step further Barowy et. al. teaches us that shepherding the crowd yields to better work of them and also to better results of the workers [21] which we already know. The statement is clearly again the crowd needs something which they can rely on it is either a person which is there for questions or it could also be the platform having organized the tasks so well that they have a self-control mechanism which helps them to lead to better results of the crowd.

Bernstein et. al. then went over to a more global question. If we have this "global brain" symbolical the understanding of things of all people combined, do we really have the environment to maximize the outcome of this? Clearly the answer is no, we need to research further in this direction to have new options to work with the crowd in a self- or guided control manner where the error of a single person is not making a big change [22]. And with respect to Barowy et. al. we know clearly that automation is the key [23] to make sure we can handle a lot of submissions of the same entry without a huge effort. Other ways the scientist would face the same problem again. Without self-control and automation, he would have to manage many people which are of all different kinds and the workload would again grow tremendously.

With the work of Minder et. al. in mind [20] De Boer and Bernstein told us how to handle a crowd task with their PPLib (PeopleLib). The PPLib acts after the five phases of Newell and Simon (Intelligence, design, choice, implementation). And then is able with CREATE and DECIDE to split and manage a crowd task that is on one hand still easy doable for the average (probably unskilled) crowdworker but with the variance of representation even lowers the risk of errors of the worker.

All this knowledge combined gives us a great view about how a crowd has to be and how it has to be managed that it reaches the most successful level of task solving which is possible. But the question is now, how to implement all this knowledge together and create something useful? And here comes finally the research of Koss and Schawinski into account. Their research about dwarf galaxies reached a big media interest since dwarf galaxies tell us a lot about the creation of the universe. They are researching deeply into the case of SDSS1133 which has been false classified first as a supernova (SN) because of a dwarf galaxy which was nearby. With this false classification in mind and the knowledge that this recoil black hole contains knowledge which gives us a deeper insight about the universe it was clear that this study needs to be developed further with the help of a crowd [27]. The "Blackhole Chaser" crowd.

3. Settings of the platform

3.1 Idea

Out of the need that with the example of SDSS1131 there are maybe a lot of other, in the same way false classified, supernovas around us. It was absolutely clear that there should be some kind solution to show this images to the masses and use them as a control and classification group. In a first draft we talk here about roughly 48'000 of such Galaxy-Images. But it is likely thinkable that there could be any number of Galaxy-Images be classified. If this platform is once established in just a small kind like Galaxy-Zoo is nowadays 48'000 classifications are absolutely no problem. And there comes the biggest pro of a crowdsourcing platform in account, not even we can classify this Galaxy-Images, we can also with a certain reliability say that a classification is trustworthy or not. And over all it should also be fun to use for the user which may or may not has a strong background in astrophysics. The technology of today makes it possible to have almost no limitations of who can see this images other than an internet connection. But until now the astrophysics field was always quite delicate since it needs a lot of understanding if you really want to know what is going on between supernovas, black holes or dwarfs. But this should not be the topic here, the topic should be to have an easy access to this pictures and try to classify them with straightforward questions. If we have the chance to establish this platform as easy as it could be we would for sure also have a lot of happy users afterwards.

Now with the example of Galaxy-Zoo it was clear that something similar is maybe useful to solve this kind of question. So the idea was clear, there should be a Web platform which is so easy to use that everyone with a computer and a steady internet connection can use it. But it should also give the worker the choice to communicate their ideas and their knowledge. The target audience should not be categorized and even an absolute non-astrophysician should be able to help in this study. But if you have a strong astrophysics background the Images presented should give you more information which is additional interesting for you as a researcher and in this case as a citizen scientist. With this Idea, a questionnaire has been done how to classify such a Galaxy-Image properly that it may or may not give information if it is a case like SDSS1131 or not at all. The tool should be two sided on one hand it should be a great experience for the user or citizen scientist which is eager to see Galaxy-Images and help to find new recoil black holes and on the other hand it should help the Group of Prof. Dr. Schawinski to have a lot of Images which are not possibly be classified with a computer. This is because for some questions you will have to spot features which are easier to solve for the human eye, like: "is there a broad point source inside the clumpy-like Galaxy?".

Following (Table 1) are the questions used in the questionnaire, all the questions have to be solved with yes or no. A no can directly lead you to a new Galaxy-Image if the general indices are not met.

Table 1. Questionnaire for "Blackhole Chaser"

Task	precondition	On "no"
Does the galaxy cloud appear clumpy?	-	Next Image
Is there a clear (often blue-ish) object visible in the picture? The colors of such so-called point sources may vary, but they should contrast their environment and often appear in blue-tones.	-	Next Image
Is the point source located within the clumpy cloud structure?	-	Next Image
Is there a point source visible in this image at the same location as in the one you have graded in the former question? For the Image of the former question, see "normal" with Tools.	Is a X-Ray Image available	-
Does the spectrum contain broad spikes that have visible white space between their lines?	Is a Spectrum available	-
Is there point distinguishable from the background at the same location as the point source identified in the former question? For the Image of the former question, see "normal" with Tools.	Is a Radio Image available	-

The tool should be two sided on one hand it should be a great experience for the user or citizen scientist which is eager to see Galaxy-Images and help to find new recoil black holes and on the other hand it should help the Group of Prof. Dr. Schawinski to have a lot of Images be classified by a lot of people. This is a huge time-win for them and the pictures shown here are not to easy be classified by a computer since it would take a long time to train this computer accurate. This is because for some questions you will have to spot features which are easier to solve for the human eye, like: "is there a broad point source inside the clumpy-like Galaxy?", for this the classification of the clumpy-like has to be done right in the first phase and a computer will not decide as a human would in this case.

3.2 Setting of the work

This Web platform should be as easy as Galaxy-Zoo and therefore it was already defined that the Platform should look very friendly and not overwhelming for the citizen scientist. Since we are talking here about a delicate topic in astrophysics we have to combine the best progress in UI-Design and still offer a lot of scientific data.

That is why the idea was near to completely extract the backend from the frontend. That even if the frontend has to be redone, the backend still can remain or vice-versa.

3.3 Backend

The Backend of this Tool contains a lot of features which are mainly used by the administrators. Therefore, they have the possibility to see which Galaxy has scored the most points (the score is given by the amount of yes answers during the questionnaire) the highest or have the possibility to see which of the Galaxies was rated as very interesting or just to see which are clumpy or not and how many people (with what percentage it is like this). Many more features are thinkable here but for the first instance of this research web platform the backend was done very straightforward. If the web platform appears to be useful (seen in pt. 4) it is a part of the future work. The backend also contains administrative features to change a description or to set up a new user, add/change coordinates and so on.

3.4 Frontend

Since the frontend should be usable for everyone, it has a lot of constraints to fulfill. It should not only be easy to use; it should also be self-explanatory on every task the citizen scientist has to complete. The self explanatory part was managed by little explanatory tour and help modals with additional explanations and examples.

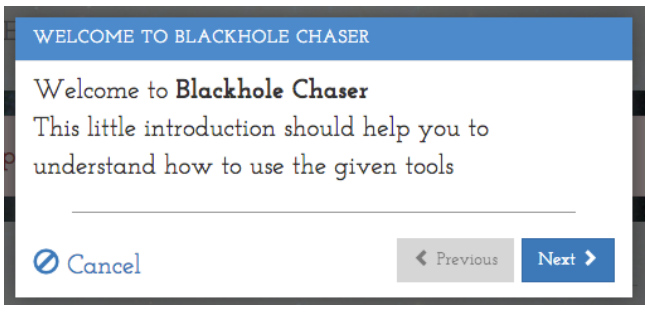

Figure 1. Start of the "Blackhole Chaser" Tour.

With this explanation the user should be ready to tackle the task which is given to him. The tasks are for this study always in order because they develop on each other (see Table 1). You will then be given a certain task to do.

Figure 2. Task description with straightforward Yes/No.

To make sure this task is good to understand we also created tooltips for the user where he finds actual examples. In the near future also Videos or interactive views could be placed here.

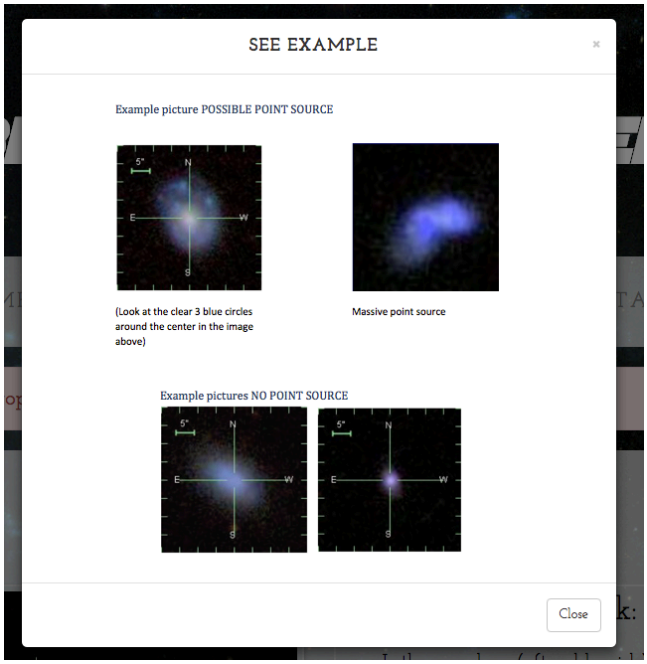

Figure 3. Tooltip as a help for the user.

For this task you may not only need to see a picture of the Galaxy, that is why it has a range of Tools have been implemented. This tools should help the user, if skilled or not to find a lot of resources about the actual Galaxy-Image.

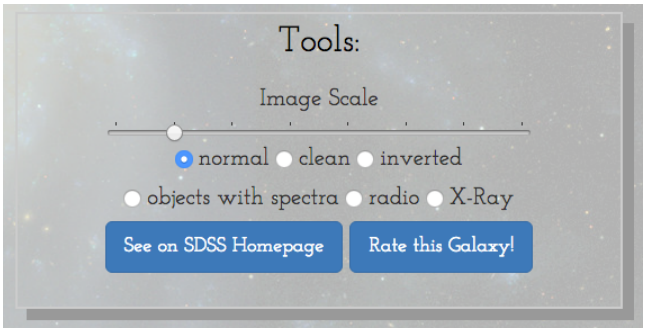

Figure 4. Tools for the "Blackhole Chaser"

With this tools you are able to set the scale of the image in arcmin by moving the slider. If you move it, immediately the next image will be shown which shows you a closer or less close image. You also have the possibility to see the Galaxy in different representations. If the user is not sure if it is really a blue-ish point source in the middle of the image it is probably useful to use an inverted view or even a clean view which lets the crosshair disappear completely. If there are still questions open, the user has always the possibility to see the picture on the SDSS Homepage which has a lot more information about each of the Galaxy-Images but is maybe not easy to understand for everyone.

Where the normal picture looks just like Figure 5:

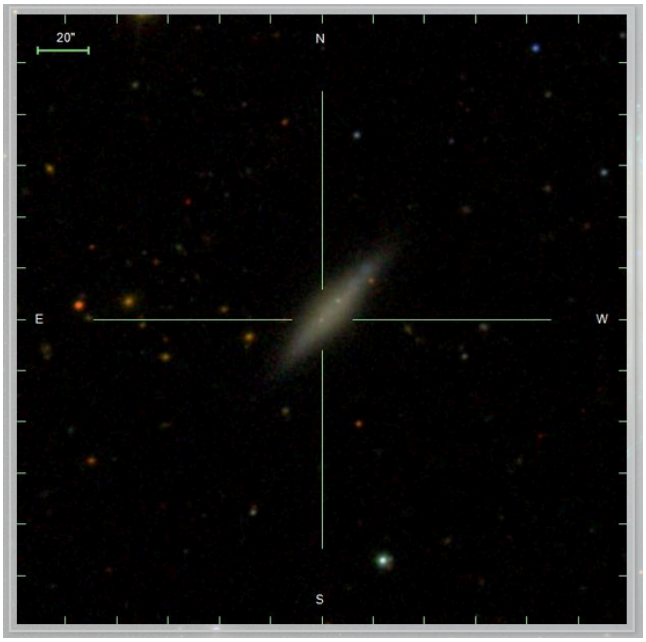

Figure 5. Representation of a normal image.

Does the radio image of this Galaxy look like this (Figure 6) which gives you the clear impression that something in the middle of this picture is worth noticing?

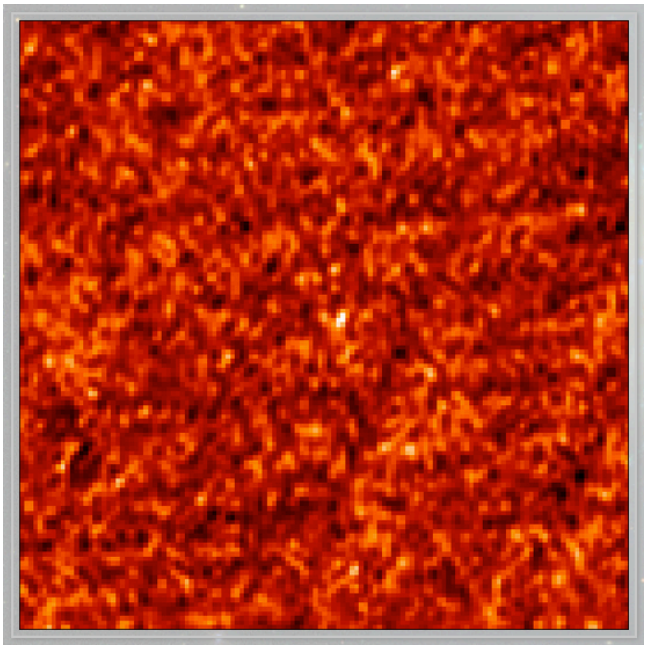

Figure 6. Radio image for the same Galaxy-Image as Figure 5.

Maybe it is even worth looking more in deep. Next we look for some spectra. If a spectrum is found, with "objects with spectra" we can locate the place of this spectra by a little red rectangular.

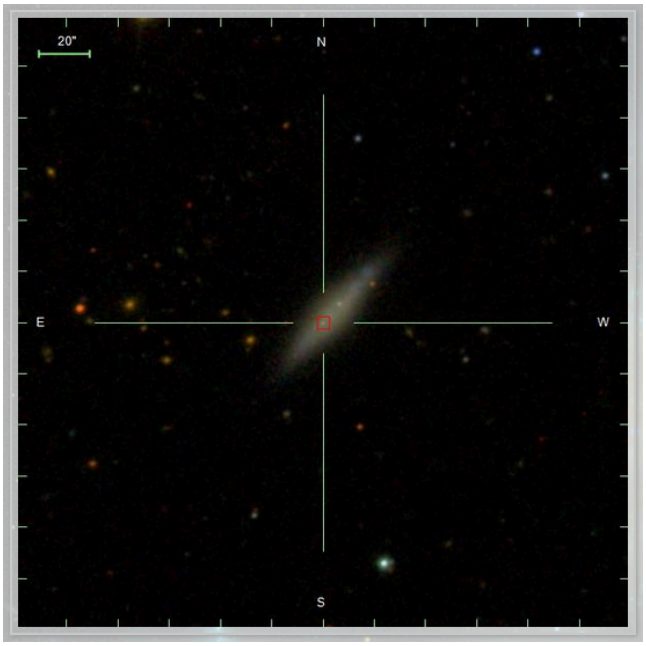

Figure 7. The same Galaxy-Image as Figure 5. as Spectra-Representation.

And then see the according spectra in detail, which contains a lot of more information for the skilled users. But for the regular users it is still an interesting image to see if the spikes are broad or not at all.

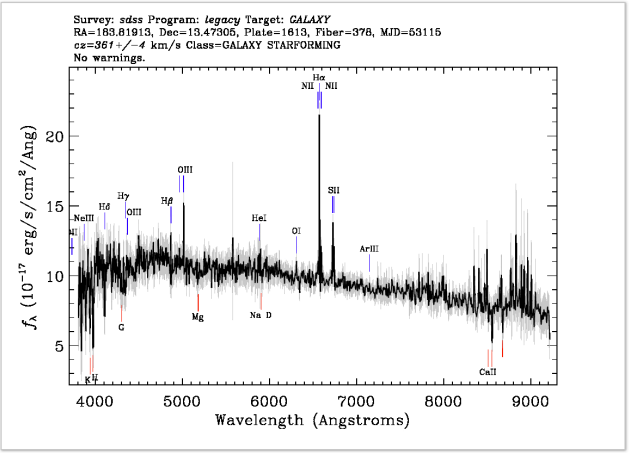

Figure 8. Spectra image for Figure 7.

To further motivate the User, we have also implemented some simple treats for the him, so that they know what they are actually looking at and how their progress in the actual test-set is

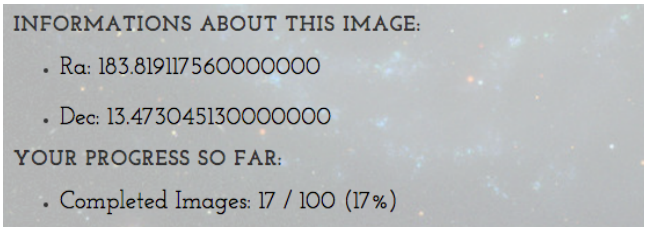

Figure 9. Status board for the actual research.

3.5 Data

The data was gathered from three different and independent data sources. We so had first to make sure how to approach this (sometimes) very old databases. But with an additional writing of 4 different download-scripts we had the possibility to gather all this data and store it on our server in 8 different scales. But we had first to make sure that the respective size in arcmin fits on all image formats.

The data was also a challenge in this case since we like to store all the image data on our servers. That is because we have no clue how long the images on their websites remain or if the servers are probably having issues because of traffic or other reasons.

Since we have in a first draft roughly 48'000 Images in 8 scales and 6 different view styles (normal, clean, inverted, X-Ray, radio and spectra) we reach closely 2 million images stored on our server which counts with some hundreds of Gigabytes in space.

4. Research Questions

4.1 Setting

To see if we are on the right track with this platform we established a little survey to handle. We needed to know how a professional (P), a trained it specialist (RG) and a regular crowdworker (CW, recruited by upWork) is actually doing on this platform. For the help of this question we would not only know if there is a huge difference between professionals and crowdworkers (and so on) we would also take into account if the individual cognitive abilities of the worker have a positive effect on the performance of the classification in general. For this reason, we use the well established Educational Testing Service (ETS) framework to see if the participant has very good cognitive skills or not and if they assist or are useless for the quality of classification. This approach was inspired by Feldman M. and his research about highly effective task assignment with ETS testing. Therefore, here are the research questions which should be solved in this paper:

RQ1: Is there a significant difference in classification of the two groups in match with a professional. If yes? Is this a general fact (over the whole group) or just on single participants?

With this question it should be shown if there is a significant difference in the group itself and if there is one to be shown clearly.

Result: To have an answer for this question we have listed all crowdworkers and people of the reference group together with their individual ETS-Score (based on our selected questions) and matched them with the first classification task they had to solve "Does this Galaxy appear clumpy?". Why have we taken this question, because only for this question we are sure that all participants have done this and reviewed this question, because if they click on "no" the search for this Galaxy is over and they will see the next Galaxy.

Table 2. Comparison of right classifications of the first task ("does this galaxy appear clumpy?") with ETS-Percentage

User	ETS-Percentage	Right Classifications (n=100)
CW1	66.23%	77
CW2*	96.10%	77
CW3	96.75%	78
CW4	83.11%	55
CW5	43.50%	30
CW6*	96.75%	78
CW7	77.27%	63
CW8	95.45%	58
CW9	37.66%	68
CW10	50.65%	58
RG1	97.40%	85
RG2	94.16%	65
RG3	96.75%	75
RG4	99.35%	72
RG5	95.45%	75
RG6	94.16%	54
RG7	99.34%	79
RG8	95.45%	66
RG9	94.80%	76
RG10	94.80%	80

*This crowdworkers were very interested in the project and asked a lot of related and well thought questions which showed a very high intrinsic motivation of this study.

If we calculate these 20 participants, we get an arithmetic mean of 68.45 correct solved tasks with a median of 73.5. If we now only calculate the crowdworkers (CW) we get an arithmetic mean of 64.2 with a median of 65.5 which is a little lower but still in a good acceptance. For the last step we have to look at the reference group of it specialists (RG) they have an arithmetic mean of 72.7 and a mean of 75. So while we are having in the reference group only people which are having at least a higher educational degree or are achieving this soon, we do not have a much higher success rate than the crowdworkers which had no requirements (still some of them are well educated). If we match this group together we see that more than 2/3 of all tasks have been successfully classified. In conclusion we can only say, what the literature told us already, the crowd liberates itself. Even if there are some over- and under-achievers the mean of the group is still acceptable. Since there is no one which achieved over 90 right classifications we sure have here space to improve the platform with more materials like videos, examples or training questions which makes it easier for the crowd to have a distinct view what is right or what is not. But still, this little test shows us that even if the crowdworkers are different educated and have a different cognitive skill level, all together they give a good overview about the classification and we can say that roughly 7/10 classifications of them are right (always in respect of the question "Does this galaxy appear clumpy").

Answer: No, there is a certain gap between this groups and the crowd manages over- and underachievers itself

RQ2: Does the cognitive skills of a worker let us know before if he will act good in comparison with a professional?

With this question we like to know if in general a more cognitive skilled worker tends to do better classifications compared to an actual professional.

Result: See Table 2, If we have someone with a really high cognitive skill, shown in our test with over 96% right answers he or she also has a good score in the classification of clumpyness (>72). While there are other good ETS-Results which were above 80% have almost no finding in coincidence. Still we have to say that all of them are for sure acceptable with mostly 2/3 of right solved classification but the span from good to moderate result is given much more than we see it for the high achievers at 96% and more.

And how are the other people which are not over 80% in their cognitive skill test. Even here we can not say that people which do not so well are not good in classifying (see CW1) and there are also example which try to show us exactly this (see CW5) but in comparison there are also participants which are breaking this rules directly (see CW9).

Well we do have to confess that a study with only 20 participants is probably not telling us enough about this occurrences, but in general we can say that if you are very strong in case of cognitive skills it is much likely that you will do very well in this classification. But even if your cognitive skill is not developed as good it is still possible to achieve good or very good results.

We have here some examples with a strong ETS-Percentage which did only moderate in classifying but we have to keep in mind by a median of 73.5 even some "bad" results are not destroying the crowd work which was very well in this case.

Answer: Yes, if there is a really high level of cognitive skill it is most likely that the classification is very good.

RQ3: Is the work of crowdworkers (paid workers, without any requirement of knowledge) usable for the classification of recoil black holes.

With this question we like to know if crowdworkers which had no requirements of skills for this study are in groups near to a professional or not. If not, we would have to keep in mind that the web platform is not yet released and it is still a beta stadium and possible misunderstandings could occur due to usability issues. But we also have to keep in mind that the crowdworkers were always allowed to ask for advice which does not contain the solution but to clarify this question.

Result: As we see in Table 2 there are certain results of crowdworkers which are very well usable. It is a true fact that the main point of these workers is to do it for money and not for the matter of the research. But still, some crowdworkers I have had to talk with were very enthusiastic about this study and went on with a lot of feedbacks and nice to haves we could in future work on. And I could clearly see, such enthusiastic crowdworkers did a fine job (See CW2 and CW6). There is a noticeable coincidence that high achievers in the ETS-Test are also more interested in the topic and not only are doing it just for the salary and in the end have also better classifying results.

As this has shown crowdworkers are not the best resource for such a platform. It is nevertheless useable for them but if there is no interest in the material or general in science the output will be only medium. So to say crowdworkers are a valuable resource for such work but they have to state their motivation. What we have proved with our question and is also shown in the results, crowdworkers with a high amount of interest in the topic are valuable for this platform and others are not bad for it since they still classify with a medium-high accuracy.

Answer: Yes, but only if the intrinsic motivations helps being exact during the classification

5. Conclusion

With the results above it has been shown that the crowd can very well manage itself also on this platform. Since the platform is definitely not yet on the level of Galaxy-Zoo we have also noticed that more example work and also a training feature would help to make the overall classifications better.

There was also a nice feature built-in in the platform where the workers could rate & comment a galaxy. Some of the workers have used this feature heavily and told us what they see and what disturbs them. This was generally only a proof of concept to move away from like buttons of Facebook and equals but it showed that it has the power to make classifications which are made, much more understandable. So if we manage it, in the future of this project to establish this feature with a well-designed backend, we could have not only the classification channel to build up tendencies where an interesting galaxy could be found or not, we could even take this in account to make the classification-score less or more worth. And we would also profit if the image is just not well shot or the light in it is disturbing, because we would see quite fast if many workers are complaining about the same pictures. It would also be an approach for machine learning to see if these comments are more positive or more negative or if they are not usable for classification.

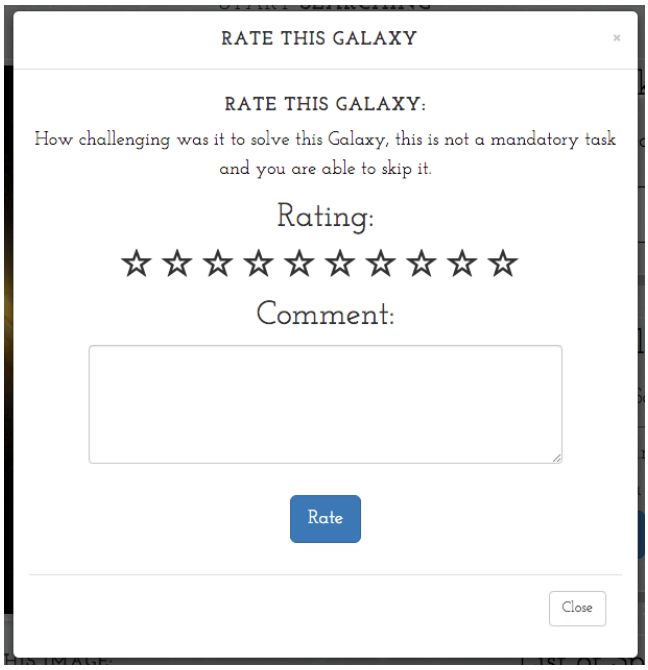

Figure 10. Rating form which can be used for each Galaxy-Image.

Here are in addition some extracts from the comments data table (sdss_id: comment)

588017702395904022: Look beautiful

Figure 11. Clean image of SDSS _ID 588017702395904022.

588017948813492321: Such dust lane, v pretty

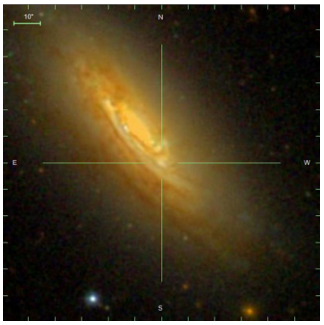

Figure 12. Crosshair view of SDSS _ID 588017948813492321.

587738947743580222: its looking like jumping!

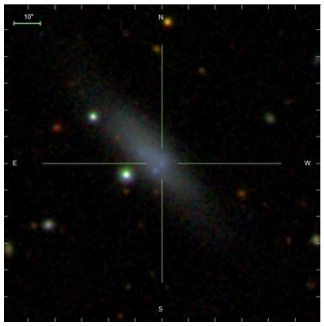

Figure 13. Crosshair view of SDSS _ID 587738947743580222.

With all this tools combined most of the participants of the first beta test had a lot of fun and were able to classify the pictures themselves.

6. Future Work

6.1 Methods

For the near future, it would also be interesting to have more established citizen scientists to test and feedback on this platform. Until now, we have not enough results to have a good idea of the view of a citizen scientist which has already worked dozens of hours at Galaxy-Zoo. It would be very interesting if he acts better than the participants here right from the beginning because he already knows how to handle such a platform.

It would also be very interesting to have more real professionals to test. Until now we only have 4 professionals where one is the "main" professional and the 3 others are references. We should also confront the main professional with some differences which all 3 other professionals have discovered to see where the misunderstanding comes from and how we could ensure it to handle such questions/misunderstandings directly by the platform so that professionals are likely to score over 90/100 classifications right.

Assisted work would also be a great opportunity to have more insight in the feelings of a worker. Over the internet it is most likely that the workers do not talk about everything which bothers him. But if we would sit next to a participant, we may have better understanding what is annoying and where he has (unspoken) troubles. We could also live talk with him about the differences he did in comparison to a professional, to see why the participant did this mistakes and to teach him how not to do it next time or even to raise a discussion with the actual professional.

Last but not least, longtime-studies would also help. Does a worker get better over time? Or does he only get better if he gets thought about a certain mistake he does. Following questions are, can this teaching be done digital or does it need somebody to teach in person in other words, by video.

6.2 Backend

Since the Backend had not really the highest amount of interest in this first prototype, it is definitely a part to work on for the future releases. Gathering the data right and with a pleasant Frontend is a good start but we would now also need to establish certain graphics in the backend which clearly show which Galaxies are marked as important or which ones are not as relevant as others. Since the subject of the broader study is still to gain knowledge about SDSS1131 alike cases, this task has to be done with a close

partnership with the actual future administrators. The people of Prof. Dr. Shawinskis Galaxy and Black Hole Astrophysics Group at the ETH Zurich will know exactly how to handle this and how to make the information best viewable for their needs.

6.3 Frontend

This first frontend works nice but also there is a lot of potential to improve. First of all, it should be completely mobile viewable. It still has some minor flaws in the layout which makes browsing on a phone or a tablet not as pleasant as it should be.

On the other side, the experiment has shown that the given explanations were not for every task enough detailed. The GUI should still deliver examples as it does right now but more alike as it is on Galaxy-Zoo right now. It should also be doable without reading any line of explanation just by having easy pictures to match (See Galaxy-Zoo as a reference).

As a last point to make this frontend even more appealing, there should be explanatory videos which show how an actual professional is doing a classification on an actual example and why he is doing it like that. So that the citizen scientist / non professional user has a good idea what to do and how.

6.4 Frontend-Features

We also should think about the development of such a citizen scientist. As soon as we have enough classification from an actual professional and some reference people we can give the citizen scientists the option to spot their mistakes or what they did wrong based on statistics. It should be kind of a “what did the professional say about this image” and it should also cover additional explanations if given to tell the citizen scientist “why it is to be classified like this”.

7. ACKNOWLEDGMENTS

My personal thanks goes to Patrick de Boer and Michael Feldman from the Dynamic and Distributed Information Systems Group (DDIS) of University of Zurich for their support and knowledge about crowdsourcing and the related material. As well as Prof. Dr. Kevin Schawinski and Andrea Scanzio from the Galaxy and Black Hole Astrophysics Group of ETH Zurich for their knowledge about recoil black holes and ways how to gather the needed galaxy image data.

8. REFERENCES

[1] Axten, N., Newell, A. and Simon, H. a. (1973), "Human Problem Solving.", Contemporary Sociology, Vol. 2 No. 2, p. 169.

[2] Malone, T.W., Crowston, K., Lee, J.L.J. and Pentland, B. (1993), "Tools for inventing organizations: toward a handbook of organizational processes", [1993] Proceedings Second Workshop on Enabling Technologies@m_Infrastructure for Collaborative Enterprises, Vol. 3, pp. 425–443.

[3] Malone, T.W., Malone, T.W., Crowston, K. and Crowston, K. (1994), "The interdisciplinary study of coordination", ACM Computing Surveys, Vol. 26 No. 1, pp. 87–119.

[4] Nau, D.S. and Nau, D.S. (1987), "Hierarchical Abstraction for Process Planning", *Second Internat. Conf. Applications of Artificial Intelligence in Engineering*.

[5] Bernstein, A. (2000), "How can cooperative work tools support dynamic group process? Bridging the specificity frontier", *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work (CSCW '00)*, pp. 279–288.

[6] Bersteins, M., Bersteins, M., Tan, D., Tan, D., Smith, G., Smith, G., Czerwinski, M., et al. (2009), "Collabio: A Game for Annotation People within Social Networks."

[7] Sheng, V.S., Provost, F. and Ipeirotis, P.G. (2008), "Get another label? improving data quality and data mining using multiple, noisy labelers", *New York*, pp. 614–622.

[8] Malone, T.W., Laubacher, R. and Dellarocas, C. (2009), "Harnessing crowds : Mapping the genome of collective intelligence", *Elements*, Vol. 1, pp. 1–20.

[9] Bernstein, M.S., Little, G., Miller, R.C., Hartmann, B., Ackerman, M.S., Karger, D.R., Crowell, D., et al. (2010), "Soylent: a word processor with a crowd inside", *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, pp. 313–322.

[10] Kietz, J.J., Serban, F., Bernstein, A. and Fischer, S. (2010), "Data mining workflow templates for intelligent discovery assistance and auto-experimentation", *Third-Generation Data Mining: Towards Service-Oriented Knowledge Discovery SoKD*, Vol. 10 No. September, pp. 1–12.

[11] Serban, F. (2010), "Auto-experimentation of KDD workflows based on ontological planning", Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 6497 LNCS No. PART 2, pp. 313–320.

[12] Bernstein, M.S., Marcus, A., Karger, D.R. and Miller, R.C. (2010), "Enhancing directed content sharing on the web", *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, pp. 971–980.

[13] Raddick, M.J., Hopkins, J., Bracey, G. and Gay, P.L. (2010), "Exploring the Motivations of Citizen Science Volunteers", No. Root.

[14] Little, G., Chilton, L.B., Goldman, M., Miller, R.C. and Link, C. (2012), "TurKit : Human Computation Algorithms on Mechanical TurKit : Human Computation Algorithms on Mechanical Turk", *New York*.

[15] Franklin, M.J., Kossmann, D., Kraska, T., Ramesh, S. and Xin, R. (2011), "CrowdDB", *Proceedings of the 2011 international conference on Management of data - SIGMOD '11*, p. 61.

[16] Ahmad, S., Battle, A., Malkani, Z. and Kamvar, S. (2011), "The Jabberwocky Programming Environment for Structured Social Computing", *Architecture*, p. 53.

[17] Colab, C. (2011), "Reducing the Load on Expert Judges in Open-Innovation Contests Preliminary Results from the Climate CoLab."

[18] Zhang, H., Law, E., Miller, R., Gajos, K., Parkes, D. and Horvitz, E. (2012), "Human computation tasks with global constraints", *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, p. 217.

[19] Weld, D.S., Adar, E., Chilton, L., Hoffmann, R., Horvitz, E., Koch, M., Landay, J., et al. (2012), "Personalized online education—a crowdsourcing challenge", *Proceedings of the 4th Human Computation Workshop (HCOMP-12)*, pp. 159–163.

[20] Minder, P. and Bernstein, A. (2012), "CrowdLang: A Programming Language for the Systematic Exploration of Human Computation Systems", *Proc. of the 4th Int'l Conf. on Social Informatics (SocInfo '12)*, pp. 124–137.

[21] Dow, S. and Kulkarni, A. (2012), "Shepherding the crowd yields better work", ... *Cooperative Work*, pp. 1013–1022.

[22] Bernstein, A., Klein, M. and Malone, T.W. (2012), "Programming the global brain", *Communications of the ACM*, Vol. 55 No. 5, p. 41.

[23] Boer, P.M.D.E. and Bernstein, A. (2014), "PPLib : Towards the Automated Generation of Crowd Computing Programs using Process Recombination and Auto- Experimentation"

[24] Feldman M. "Cognition-based Task Routing : Towards Highly-Effective Task-Assignments in Crowdsourcing Settings", pp. 1–10.

[25] Zhang, H., Chen, Y., An, F., Francisco, S., Ahn, L. Von, Liem, B. and Chen, Y. (2015), "An Iterative Dual Pathway Structure for Speech-to-Text Transcription The Harvard community has made this article openly available . Please share how this access benefits you . Your story matters . Citation Human Computation : Papers from the AAAI Workshop."

[26] Bernstein, A. (n.d.). "The Global Brain Semantic Web Interleaving Human-Machine Knowledge and Computation ∗", pp. 1–6.

[27] Koss, M., Blecha, L., Mushotzky, R., Hung, C.L., Veilleux, S., Trakhtenbrot, B., Schawinski, K., et al. (2014), "SDSS1133: An Unusually Persistent Transient in a Nearby Dwarf Galaxy", Vol. 527 No. January, pp. 515–527.