New approach to interactive terminologies discussion

drs. M.J. (Marielle) Timmer, dr. E.J. (Julian) Bianco Martinez, J.J. (Jim) Hoekstra MSc,
prof.dr.ir. J.L. (Jan) Top

WAGENINGEN
UNIVERSITY & RESEARCH

New approach to interactive terminologies discussion

Authors: drs. M.J. (Marielle) Timmer, dr. E.J. (Julian) Bianco Martinez, J.J. (Jim) Hoekstra MSc, prof.dr.ir. J.L. (Jan) Top

Public

Report 2232

WAGENINGEN
UNIVERSITY & RESEARCH

The research that is documented in this report was conducted in an objective way by researchers who act impartial with respect to the client(s) and sponsor(s). This report can be downloaded for free at https://doi.org/10.18174/559037 or at www.wur.eu/wfbr (under publications).

Contents

Summary

This report describes the result of the third year research of the Terminology Project within the Multiple Scale and Extreme Events motive of the Food Systems program of Wageningen University & Research. The aim of this project is to create support for researchers from different teams and disciplines to quickly create a shared understanding of the terms to be used in their collaborative project or task.

Given the diversity of experts and stakeholders in research areas such as Food Systems, confusion and misunderstanding of the terminology used can easily arise. The interpretation of terms often remains implicit, which may cause problems at a later stage. Our assumption is that an interactive visualization of the terminology will be helpful. Rather than trying to set up full definitions for the terms used, simple 'triples' can express the meaning of these terms. Triples are elementary phrases in the form of subject-predicate-object (e.g. 'delta – has issue – flooding'). Visualization of such association with a simple tool can support research teams in quickly creating some level of common understanding and demarcation of terms within a project or task.
We also assume that associated terms could help to define a term and that it will be much faster and fun if these associated terms are generated automatically by the tool. For this we can build on results in the field of semantic association computation.

With TALK, Team Associations for Linking Knowledge, we successfully created a game and supporting tool that already to some extent meets these requirements. The teams were able to clarify a set of core terms without the need for full definitions. The visualization and interactive use of the tool effectively assisted the discussion and the participants enjoyed using the tool. Automatic creation of associations does reduce the workload for generic terms, but not so much for specific terms.

The evaluation sessions with TALK lead to a number of proposals for further improvement of the game and tool, and on how the outcomes can be used further. For example, it could be based on a more specific repository for generating more relevant associations, such the Wageningen Library. Also, storage of the graphs and using the results for creating or extending controlled vocabularies could be a useful next step.

1   Introduction

This report describes the result of the third year of research of the Terminology Project within the Multiple Scale and Extreme Events motive of the Food Systems program of Wageningen University & Research. The aim of this project is to create support for researchers from different teams and disciplines to quickly create a shared understanding of the terms used in their collaborative project or task. We submit that a brief session supported by a simple tool can already be effective.

Given the diversity of experts and stakeholders in research areas such as Food Systems, confusion and misunderstanding of the terminology used can easily arise. The interpretation of terms often remains implicit, which may cause problems at a later stage. Moreover, the scope of the considered task can also be conceived differently by different players, without being aware of it. It is clearly not feasible to create a single, universally valid language for all circumstances. However, it is possible to at least briefly review the main terminology for a specific research action. The question is how to facilitate this process such to make it effective and efficient.

Our assumption is that an interactive visualization of the terminology can be helpful. Rather than trying to set up full definitions for the terms used – which can easily result in endless debates – simple 'triples' can express the meaning of these terms. Triples are elementary phrases in the form of subject-predicate-object (e.g. 'delta – has issue – flooding'). By providing a set of possible related terms (associations) for each base term, clarifications can be constructed without the need for full definitions. Visualization of such association with a simple tool can support research teams in quickly creating some level of common understanding and demarcation of terms within a project or task.

We also assume that automatic generation of the associated terms, given some term to define, will make the tool more effective. For this we can build on results in the field of semantic association computation. For an overview of this field we refer to [1]. In our project, we use so-called Word2vec models of selected repositories for generating associations, but this is only one way to go.

In 2020, our first plan was to start the creation of a shared vocabulary by asking researchers in one of the MSX motives to collect a small set of terms, provide definitions for these terms and specify related terms. However, it was difficult to motivate researchers to give such input without seeing the immediate benefit. This motivated us to develop a new approach, in which we take into account the question how to motivate people to participate.

The first important innovation was that we created an algorithm to automatically retrieve term associations from the web, thus giving the discussion between the experts a head start and reducing their workload. Next, we developed an interactive tool that visualizes these associations as a graph. We also developed a process or 'game', to support the expert discussions. In this way we added a fun element to the expert discussions that also serves to motivate the participants. We develop this game and tool in an agile way by maintaining a product vision document to agree on the scoping and by adding functionality in sprints.

2   Related tools

Several digital tools are available either commercially or open source to support discussions, brainstorms, creative sessions, etc. Most of these are general drawing tools, or ways to create and organise notes. They often facilitate drawing graphs (min maps), pictures (icons) or (sticky) notes. A blackboard architecture allows multiple users to work on the same document, which can be stored for later inspection or reuse. An important aspect for such tools is that they are easy, fast and fun to use.

One of the goals of such tools is to clarify the meaning and use of concepts in a group of people. An obvious way of presenting the meaning of words is by giving definitions in one or more sentences. However, this is cumbersome and can lead to lengthy discussions. It is more practical to explain terms by their associations with other terms. A number of tools and games already work with associations.

- Text editors such as MSWord, Google Docs and Latex provide synonyms for words used in a text to assist authors in expressing themselves.
- Codenames is a card game in which wo teams compete by giving one-word clues that can point to multiple associated words on the board. The other players on the team attempt to guess their team's words while avoiding the words of the other team[1].
- In speech therapy for children 'Sort-A-Scene Boom Cards' help students to learn about categories, associations, and semantic relationships as they sort items into the appropriate scene[2].
- Play-a-LOD is a card game in which the players build 'triples', i.e. subject-predicate-object relations, in order to represent data in a graph of semantic associations[3].
- The game 'Word Association' is a simple game based on free associations. One of the players proposes some word. The first participant lets the second know this word, usually on a sheet of paper. The second participant writes down an associated term. Each player sees only the word written by the previous player. The process is repeated several times by turns, and then one of the participants reads out the resulting list of words. There will be surprises, of course, because each person has its individual associations for the given word[4].
- 'wordassociation.org' is a simple game to chain associations, based on one of the largest database of word associations[5]

For creating such tools several collections of words with their associations are available, such as
- 'RelatedWords' helps you find words that are related to a specific word or phrase. '*It runs on several different algorithms which compete to get their results higher in the list. One such algorithm uses word embedding to convert words into many dimensional vectors which represent their meanings. The vectors of the words in your query are compared to a huge database of pre-computed vectors to find similar words. Another algorithm crawls through Concept Net to find words which have some meaningful relationship with your query. These algorithms, and several more, are what allows Related Words to give you... related words - rather than just direct synonyms*'[6] .
- 'Wordassociations.net' provides associations and descriptions for a word given by the user. An API is available to access to the Word Associations Network online. It currently supports seven languages and allows finding associations for a separate word or a whole phrase[7].
  Visuwords™ is a tool that presents a visual view of qualified associations between words (a semantic graph).

---

[1] https://en.wikipedia.org/wiki/Codenames_(board_game)
[2] https://www.teacherspayteachers.com/Product/Semantic-Associations-Sort-A-Scene-Boom-Cards-Distance-Learning-3653029?st=d12c299b2a4699477dbeb1660e21ba10
[3] https://www.pldn.nl/wiki/Play-a-LOD
[4] https://wordassociations.net/en/games
[5] http://wordassociation.org/about/
[6] https://relatedwords.org/
[7] https://wordassociations.net/en/

However, none of these tools has been developed for supporting collaboration in research teams. Therefore we have designed a new game, partly based on ideas already available in the listed tools and partly based on our own experience with semantic modelling and machine learning.

3 Design and development of the tool and game

The objective for developing the TALK (Team Associations for Linking Knowledge) game and tool is to create a quick and pleasant method to get consensus on the main terms in research project teams that are at the point of starting a new collaboration. This method may also help to refine the precise goal of that research task.

For this, we make the following assumptions:

- The participants (about 4-8 people) work together in a game setting, supported by a digital tool.
- The tool should be easy, fast and fun to use. This suggests that we need automatically suggested associations, a visual representation and simple editing functionality.
- The game and tool should prevent 'a priori funnelling', i.e., pushing the participants in a certain direction. However, at the same time suggestions that are completely out of scope should be kept to a minimum.

Given this overall goal and these assumptions, we employed an agile (iterative) development process. In this process, we (re)defined the rules of the game, the way to generate associations, visual representation method end editing functions.

3.1 Design and development of the tool

3.1.1 Sprint 1: Enter a word, find associated words and display them in a word cloud

As a starting point we have the players individually select a few key terms for the task at hand and let the tool generate about five associated terms. These are represented as a cloud with a core term and multiple associated terms. At first, we set the rule that one person would first retrieve the associated terms for his own core terms and add or delete terms to his own clouds. However, during the first sprint we noticed that it can be very useful to discuss the initial clouds terms as a team from the beginning and have the 'active player' explain the edits he or she is making.

For generating associations the question is: how to automatically find associated terms for a given word? For this purpose we use word cloud, generated by *word2vec* algorithms [2] applied to a large repository of text documents. To prevent funnelling into specific disciplinary funnelling we selected the most generic repositories we could find, i.e., Twitter and Google. The associated terms generated with Twitter gave more meaningful words than the words generated by Google, so we continued with the first. We might use thesauri in a later stage, since these might also be biased for a specific discipline.

For the visual representation we decided that an overview showing all clouds simultaneously would be overwhelming and would lead to loss of focus. We therefore show one word with association at a time. Showing seven associated words for one word is a good amount. The initial editor allowed a player to enter a core term and view the generated cloud.

At this point we evaluated the tool in our own team. We already noticed the fun factor of it.

## 3.1.2   Sprint 2: Add and remove terms from the cloud

A first technical improvement for the tool would be for players to be able to remove and add associations in the generated clouds, as these should be optimised for their interpretation of the core terms in the given research context.

For the visual representation we decided to

- give the core term a different colour, such that the difference between the core concept and the associated words is clear,
- Keep the associated words in a fixed position. Currently, when a word is removed the other words are shifting to a new place. In this way it is not clear which word has been removed (see Figure 1 and   Figure 2)
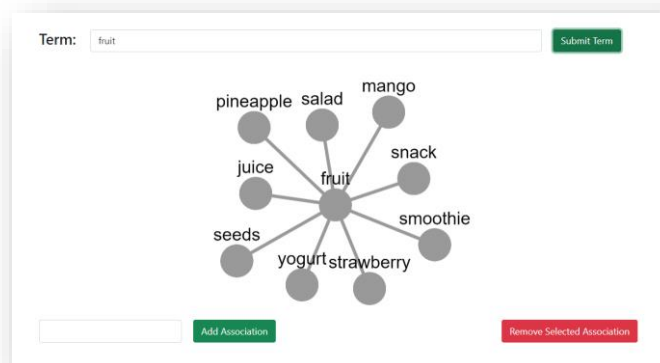
Figure 1   Word cloud before removing 'seeds'
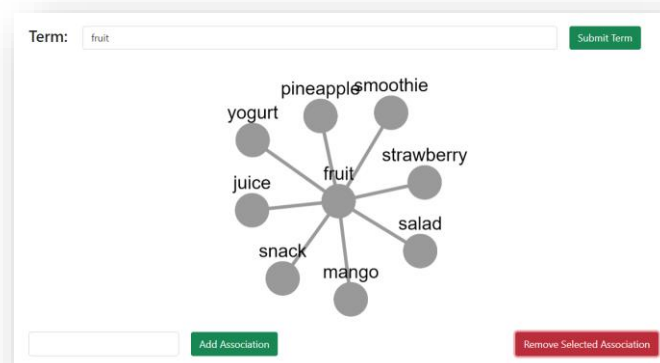
Figure 2   After removing 'seeds'

We discussed the key rules and tool functionality again. We wondered whether it would be possible to also support the selection of the core terms. It would be an added value if the tool would find, with minimal effort from the participants, the words that can cause miscommunication and misunderstanding within the group. For example, the term 'delta' is associated with 'river' by a geographer, but to 'difference' by a mathematician. In what technical ways can we help a diverse group of people to find that words causing potential miscommunication?

One way to approach this would be the following:

1. Let all the participants come up with a number of core terms
2. Select clouds with the same core term.
3. Count how often the same term is in the found associated word clouds (or another similarity measure).
4. If more people have the same word as core concept have a look at the associated terms. If there are differences, use these words in the discussion

Another idea is based on the fact that the participants come from different backgrounds. Use a set of main terms from the respective disciplines in combination with terms from the project description to find terms that (do not) fit both contexts.
We decided to give this issue of selecting core teams not a high priority in this sprint.

 Other points that were discussed in this phase:

- As questioned before: should the game be played individually or with the group? Which approach motivates the most? Selection of the core terms would probably be done individually up front. Next the editing of the clouds could be done in the team by taking turns.
- Regenerate the graph after the first cleaning action to combine clouds. What would be the added value of this?
- There is a need to be able to remove multiple associations at once.

3.1.3 Sprint 3: How to grow the association

We worked on the last requirement specified in the previous sprint. With *Ctrl+left mouse* click you can select multiple associations in a cloud and remove them all at once with the button 'remove selected associations'
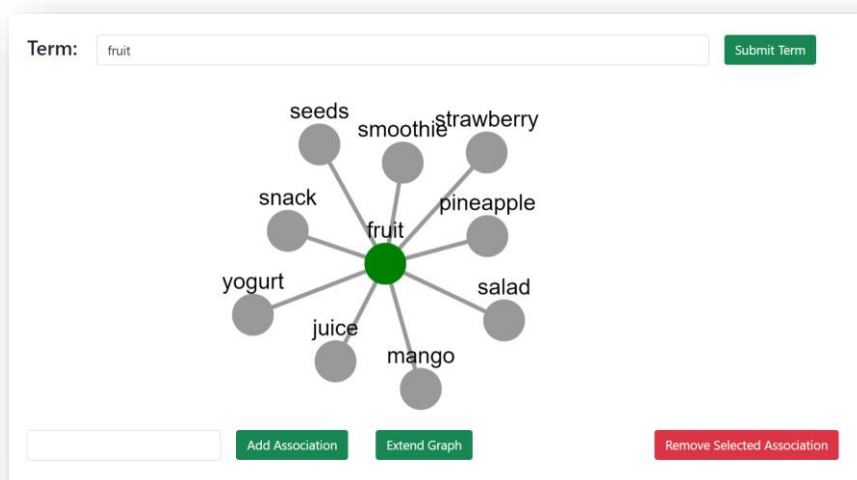
Figure 3 Removing multiple associated terms

3.1.4 Sprint 4: Most important changes to use the tool in the test with the project team

In order to prepare the tool for the first test by other users we covered the following issues:

- In the Twitter dataset, all the terms are lower case. When users typed a word using an upper case the tool gave no result. In the background a search term with upper case is now changed to lower case to solve this problem.
- The tool generates more nodes with each request. This eventually results in a large cloud that cannot be read. We decided to set the maximum number of associated terms to nine.
- When you add multiple words they all appear in the same place in the graph and overlap each other. We solved this by positioning the newly added terms between the other nodes

3.2 The rules of the game

After sprint 3 we had developed the tool far enough to be able to test it with our own Food Informatics team. Next, we wrote the manual which was necessary for the follow-up evaluations, for example with researchers from the 'Multiple Scale'-project. This section summarizes the rules of the game as presented in the manual.

3.2.1 Way of working

The goal of 'game' is to create a shared understanding of the terminology used in collaborative research projects. In particular in multidisciplinary teams the terms used in meetings and reports can easily be understood in different ways by different actors. TALK aims to reduce misunderstandings with respect to terminology and objectives of a project or task. For example, the term 'delta' can be understood as the mouth of a river, a Greek letter or an airline. And if it is understood as part of river, it can still be considered from a geographic, social, economic perspective. Which aspects do we concentrate in our project team? We are often not aware of these different interpretations and expectations, causing unintended misunderstandings.

The meaning of terms can easily be made explicit by listing other terms that we associate with them. For example, the term 'delta' can be associated with 'river' and 'wetland', but also with 'gamma' or 'aircraft'. Or more specifically, with 'flooding', 'agriculture', 'food', etc.
The idea behind TALK is to quickly create associations for a selected set of terms and seek for consensus of which interpretation is intended for a specific project or task. This can be done in short sessions at the beginning or a later stage of the project. The TALK tool helps during these session in generating, visualizing and editing these associations. However, the main discussion between the actors on the selection of terms and how they should be interpreted is the main activity, i.e. to TALK.

The TALK method consists of three steps:

1. Selection: Individually select ~5-10 core terms that you consider important for your project, and which should be properly understood by all team members.
2. Divergence: Create your individual association graph using the TALK tool
3. Convergence: Explain and discuss the created association graphs and adapt where necessary to reflect the shared understanding of terms and objectives in the project.
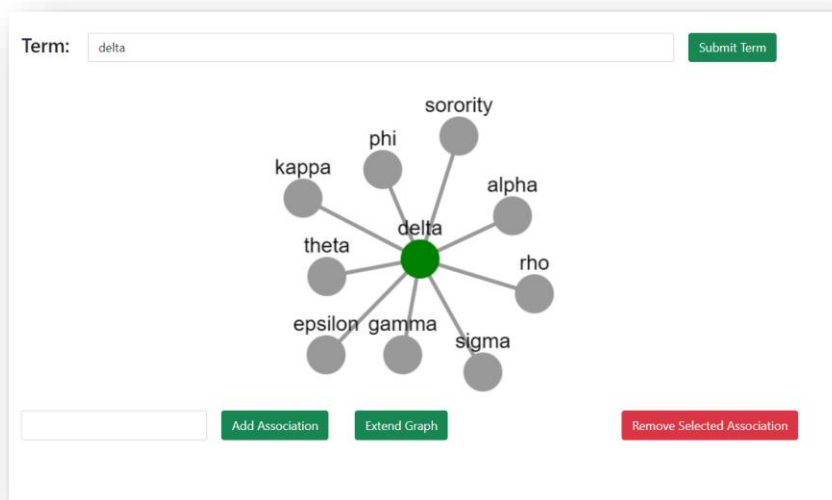
Tool
- Access the TALK tool at http://scotty.wurnet.nl:8083/msx/. Note that this URI is only accessible via the WUR network (or VPN).
- Enter a selected term and click Submit Term.

- A small graph will show, with selected term in the middle and its associations linked to it. This graph is generated automatically from a large body of texts.

- Remove terms that are not applicable to your case by selecting them and applying Remove Selected Associations. You can remove multiple terms at once by selecting them with Ctrl-click and then by pressing Remove Selected Associations
- Add terms manually to express the associations explain how the selected term should be used in this project, or which aspects of it are important. Write the term in the text field at the bottom of the page and press Add Association.
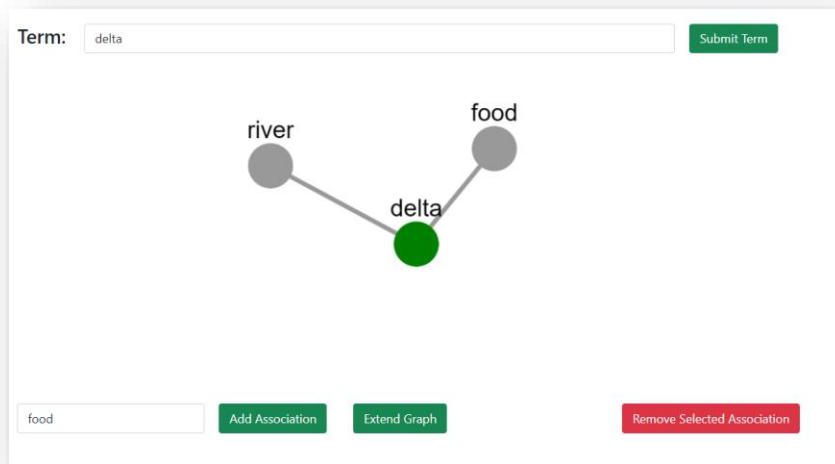
- If you wish to get more suggested associations, you can press Extend Graph. This will retrieve terms that are linked to the current word cloud.
- Repeat the steps until you think the cloud properly represents the core term. You can apply any of the above steps at any time, the order is not fixed.
- Currently we support terms consisting of single words only.
- Try to keep the size of the graph small, max ten associations.
- You are done if some consensus has been arranged for all selected core terms, or if for some terms the discussed on that has been deferred.
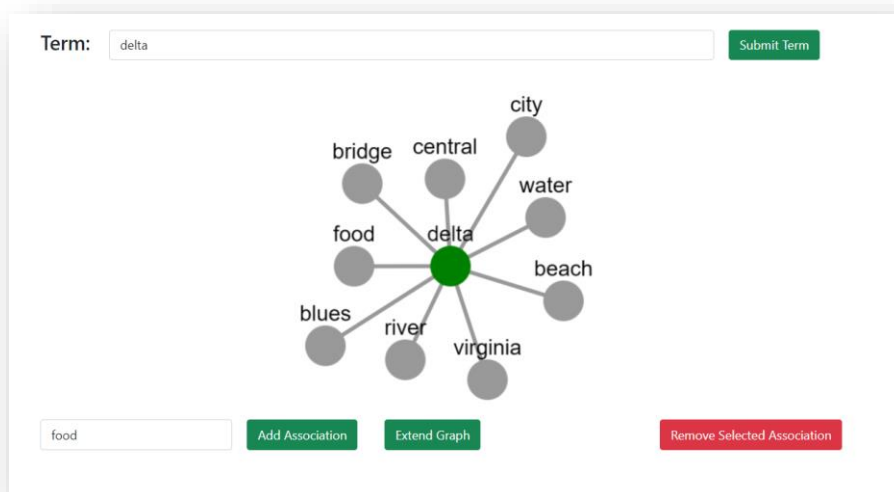
- You can make screen shots of the created associations and keep them for future reference. At some point you may choose to do a follow-up session with the same or other terms.

4 Experience from sessions with WUR researchers

This chapter describes the experiences obtained in sessions with the 'Food Informatics' and the 'Deltas under pressure' teams.

4.1 Observations Food Informatics session

The group was divided in Team 1 and Team 2, each consisting of about five people.

4.1.1 Team 1

- Not everyone choses words that are related to Food Informatics subjects. It should be made clear that the associated words should be constrained to the chosen context.
- The process of explaining the associations for a term could trigger new ideas for associations, or doubt about the associations that had already been chosen.
- Sometimes when you automatically extend the graph it happens that previously removed words return in the graph.
- If one of the associated terms gives a lot of discussion this could be the next term to be discussed
- Misspelled words don't give any results.
- The process took more time than planned, we finished the graphs of only two people in the given time.

4.1.2 Team 2

- Should everybody use the same (natural) language?
- The current tool is case-sensitive[8].
- Not all words are in the vocabulary, how to deal with this?
- Instructions on how to store the graphs resulting from the individual part of the process should be very clear, so that all participants use the same method.
- Very similar words occur in graph (e.g. 'increasing', 'increased', and 'increase'), which perhaps should be avoided.
- How critical should people be when keeping or removing associations? Is 'vaguely related given a context' enough or should everybody be able to clearly motivate why this association is important to describe the meaning?
- It should be made more clear how to use the multiselect feature to remove multiple nodes at once.

4.1.3 Evaluation with the entire group

- The 'extend' feature is nice because it triggers new associations to things you didn't think of before.
- When you add multiple words they all appear in the same place in the graph and overlap each other[9].
- Renaming of terms is not possible.
- The associations that come up in the graph are quite different from what was meant by the player.

---

[8] This has been solved in the current version.
[9] This has been solved in the current version

- If upper case is used the application gives no result [10].
- What do people want to do with the result? Do we need to store the resulting graph?
- The amount of words shown in the graph is good (it should definitely not be more).
- Suggestion: rank the results based on the distance between the association terms and the core term, and use a slider to select how many associations you want to show.
- Associations in the graph can also have multiple meanings themselves (suggestion: add a description for these words to explain the meaning)
- Perhaps we should keep track of associations that have been removed, such that they are not regenerated later.

4.2 Observations 'Deltas under pressure' session

4.2.1 Team 1

- Using composite terms (noun phrases) as core terms is not fully supported. The terms 'food system' and 'value creation' do not generate any associations, whereas 'land use', 'decision making' and 'food production' do.
- The tool does not give associations for words containing hyphens.
- It takes about ten minutes to find suitable associations for three words (per person).
- In one case thirteen terms were used to explain a core term
- A user wanted to continue to specify an associated term by using it as a core term in a new cloud. For example 'food' (what kind of food?), or water.
- A user feels that the action 'extend graph' messes up the graph, after it had been put into order by her.
- The user expects something to happen when the 'extend graph' button has been pressed. However, sometimes nothing happens.
- Sometimes it is not possible to add a term. The message 'updating' is shown.

4.2.2 Team 2

- Someone questions whether it really helps to clarify a term by adding (more) associations, because each association can give more discussion.
- How specific should the associations be?
- Some players want to add terms to associated terms.
- The suggested associations are not always relevant when the core term is specific.
- Also composite terms are added to the graph as associations.
- The group concludes that the more specific the core term is, the better the tool works.
- One of the users said 'Whoopi' while giving a term another position, he seems to enjoy himself using the tool.
- They agree on a term, but forgot to register.

4.2.3 Evaluation with the entire group

- Facilitate the use of composite core terms (noun phrases).
- The tool helps in the discussion, to diverge, converge, and clarify.
- The automatically generated suggestions were not always very helpful. Sometimes they removed most of them.
- If a term is general it is more likely that the associations are helpful than with a more specific term.
- You want to link further, use an association as core term and get associations for these.

---

[10] This has been solved in the current version

- What is the advantage of this tool compared to a mind map?
  - It is automatic and easy to arrange it in a nice way.
  - You have to make a choice, your starting point is important, so that is interesting.
  - The purpose of the tool is to describe the meaning of the words selected as core terms. If you want to describe one of the associations of the word, then this could be selected as a new core term.
  - I prefer a mind map when I start a project to explore the topic, as you can show how the terms are related. And during the making of the mind map you can discuss the meaning of the word used as well.
- Playing around with the tool is nice.
- Spending fifteen minutes per term is okay, otherwise it gets too long.
- Start individually with the same set of words to see what the interpretation is for that word and then have the discussion afterwards.
- You have to get to know the tool in the beginning, it was useful that we got the explanation on beforehand.
- Explaining to each other how you get to the terms is also interesting
- They had expected that we had used their documents in some way. These documents gave already guidance for the discussion. Now it is a loss of time because there is already an agreement on these documents.
- The game can help a new team to agree on terms and their meaning.
- When the team agrees already, the process can be faster.
- The tool can be used for monitoring whether you are on the right track in your project.
- They like to have the graphs stored, so they can come back and reuse them.

5   Conclusions

The aim of this project is to create support for researchers from different teams and disciplines to quickly create a shared understanding of the terms used in their collaborative project or task. We originally defined the following criteria for success in the product vision document:

- The method helps users to participate in the discussion with minimal effort on generating the associated terms.
- The users can discuss and work together on the terms and the associations help to get a common understanding of the used vocabulary.
- The participants enjoy the usage of the tool.

With TALK we successfully created a game and supporting tool that already to some extent meets these requirements.

The teams were able to clarify a set of core terms without the need for full definitions. The visualization and interactive use of the tool effectively assisted the discussion and the participants enjoyed using the tool. Automatic creation of associations does reduce the workload for generic terms, but not so much for specific terms.

For further development the following proposals were made:

- Use a more specific repository for generating more relevant associations? For this idea we contacted the Wageningen Library, to get access to documents that relate to the entire Wageningen research community. They show much interest in our work, as the resulting models may be useful in for example annotating documents (for improved search facilities).
- Provide support to continue the association process from the associations selected before. For example, if the term 'delta' has been associated with 'water', the latter terms can be taken as a new core term for which another cloud is created.
- Support for noun phrases is needed.
- Storage of the graphs and using the results for creating or extending controlled vocabularies, for example for search purposes.

Literature

[1] Shahida Jabeen, Xiaoying Gao, Peter Andreae, Springer Nature B.V. 2019. Semantic associations computation: a comprehensive survey, Artificial Intelligence Review (2020) 53:3849–3899. DOI: 10.1007/s10462-019-09781-w

[2] Mikolov, T., Chen, K., Corrado, G. & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. CoRR, abs/1301.3781

To explore
the potential
of nature to
improve the
quality of life

The mission of Wageningen University and Research is "To explore the potential of nature to improve the quality of life". Under the banner Wageningen University & Research, Wageningen University and the specialised research institutes of the Wageningen Research Foundation have joined forces in contributing to finding solutions to important questions in the domain of healthy food and living environment. With its roughly 30 branches, 5,000 employees and 10,000 students, Wageningen University & Research is one of the leading organisations in its domain. The unique Wageningen approach lies in its integrated approach to issues and the collaboration between different disciplines.