Robot-author

Laszlo Pitlik (sen), Laszlo Pitlik (jun), Matyas Pitlik, Marcell Pitlik (MY-X team)

Abstract: The paper demonstrates a quasi general approach how articles/studies/papers can be derived based on robot activities and available log-data in form of an object-attribute-matrix (OAM). The paper is a part of the learning material for international courses focusing on keywords like: Service Science & Knowledge Economy: Research Methods, Quality, Innovation Policies and Tools, Advanced Service Design & Management – but the general potential could also be interpreted in case of quasi arbitrary phenomena. The logical pattern about creating studies described in this paper could be used by teams of Students in order to realize articles within shortest time in a successful way. The starting point should always be an available OAM. For each OAM it is possible to choose at least 3 different interpretation ways: it is possible to analyse data compared to theoretical expectations (c.f. term creation, ranking). It is also possible to fine tune relationships between variables (c.f. production functions, simulations). The third way is: to explore system pattern in case of quasi totally unknown systems. The necessary analytical potential can be involved through online engines (like MY-X FREE – based on similarity analyses) and/or Solver-like modules (c.f. Excel, Google spreadsheets). This kind of engines can be used without any deep mathematical backgrounds. The interpretation of the calculated/modelled results could also be offered in form of online engines, but here and now, it is the task of human beings. Data-visualisation effects can be generated based on coloured backgrounds of cells in a matrix. The goals of an analysis can be defined by a human being or the robots can derive the most relevant model-type and so the most appropriate goal based on text mining where texts of human beings will be analysed.

Keywords: similarity analysis, automation, quality assurance, effectivity, efficiency, generality

# Introduction

This paper is the newest part of the series about experiences of the QuILT-based education processes. Previous articles and their annexes can be downloaded here:

1. <https://miau.my-x.hu/miau/quilt/Definitions_of_knowledge.docx> + annexes like:
   * <https://miau.my-x.hu/miau/quilt/demo_questions_to_important_messages.docx>
   * <https://miau.my-x.hu/mediawiki/index.php/QuILT-IK045-Diary>
   * <https://miau.my-x.hu/mediawiki/index.php/Vita:QuILT-IK045-Diary>
   * <https://miau.my-x.hu/mediawiki/index.php/QuILT-IK059-Diary>
   * <https://miau.my-x.hu/mediawiki/index.php/Vita:QuILT-IK059-Diary>
2. <https://miau.my-x.hu/miau/quilt/reality_driven_education.docx> + annexes like:
   * <https://miau.my-x.hu/miau/quilt/chained-translations-legal-slang.docx>
   * <https://miau.my-x.hu/miau/quilt/demo_chained_translations.docx>
   * <https://miau.my-x.hu/miau/quilt/demos_chained_translations.docx>
   * <https://miau.my-x.hu/miau/quilt/forum_details.docx>
   * <https://miau.my-x.hu/mediawiki/index.php/QuILT-IK057-Diary>
   * <https://miau.my-x.hu/mediawiki/index.php/Vita:QuILT-IK057-Diary>
3. <https://miau.my-x.hu/miau/quilt/Exercises_for_critical_thinking_and_doing.docx>

1. <https://miau.my-x.hu/miau/quilt/st1_all.docx>
2. <https://miau.my-x.hu/miau/quilt/20Q.docx>
3. <https://miau.my-x.hu/miau/quilt/GDP_final_en.doc>
4. <https://miau.my-x.hu/miau/quilt/st2_all.docx>
5. <https://miau.my-x.hu/miau/quilt/harmony.docx>
6. <https://miau.my-x.hu/miau/quilt/safety-index.docx>
7. <https://miau.my-x.hu/miau/quilt/20q_based_fingerprints_of_words.docx>
8. <https://miau.my-x.hu/miau/quilt/alternative_evaluations.docx>
9. (<https://miau.my-x.hu/miau/quilt/robot_author.docx>)
10. ((the next item will demonstrate the experiences of the paper-creating experiments))

Parallel, there are a lot of spreadsheets supporting the needs for details: <https://miau.my-x.hu/miau/quilt/?C=M;O=D>

Based on Knuth’s definition: knowledge/science is what can be transformed into source code – each other human activity is a kind of artistic performance. Therefore, the creating process of publications should also be transformed into source code. The Turing-Test could “measure” whether a paper created by a robot seems to be at least so good like a paper created by a human.

The first attempts to use robots for content development was already 2010 (<https://miau.my-x.hu/myx-free/index_fifawc2010.php3?x=soccer_news> - <https://miau.my-x.hu/myx-free/nos/>) where the objective was to create sport news based on data (sport statistics) and analytical capacities of online robot engines.

The last impulse of the decade-long experiment was the competition of CLAAS (see Unitech Competition 2019 - <https://www.unitechcompetition.com/eredmenyek>) where the competition had 2 parts – a classic simulation task and a classic innovation challenge. The innovation challenge can be interpreted as a kind of creation of a publication.

The winner of the recent competition involved the same consultation potential like the Students having the task of creating publications based on OAMs and robot engines (see: <https://miau.my-x.hu/myx-free/coco/index.html>).

The title of winner’s idea was: “Business potential of log-based services about machines in the portfolio of CLAAS” (<https://miau.my-x.hu/miau/248/claas.ppt>). It means: the best team had the assumption that a lot of OAMs could be queried at once being capable to support decision making processes where the online similarity analysis is responsible for the analytical potential as such. With other words: the second part of the competition could also be generated by a robot.



Figure Nr1: Aggregated results (source: <https://www.unitechcompetition.com/eredmenyek>)

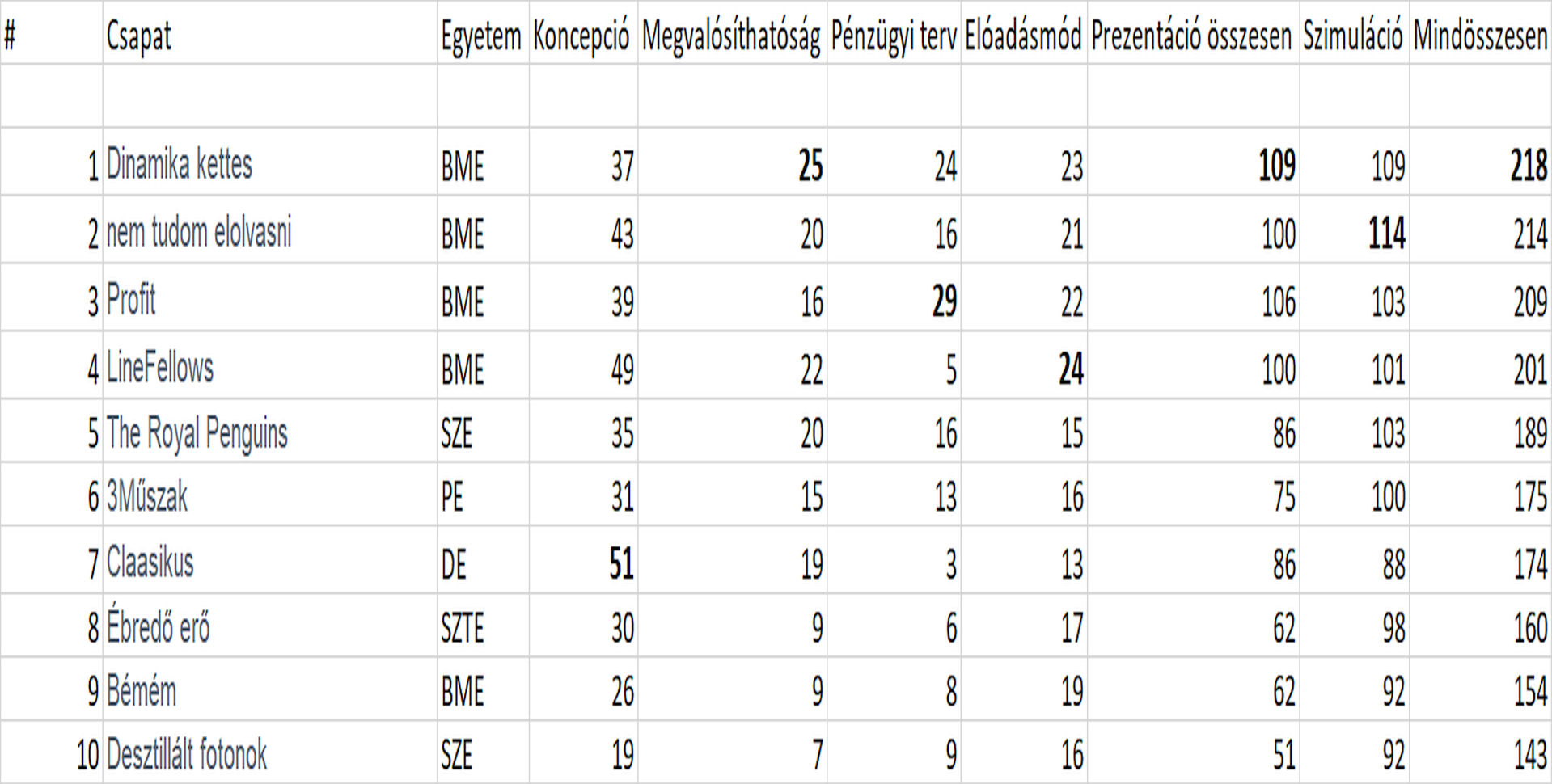


Figure Nr2: Detailed results of the presentation (source: <https://www.unitechcompetition.com/data/2019/UniTech_2019_final.xlsx>)

Figure Nr1 demonstrates that the winner became a winner through the presentation (second task). Figure Nr2 shows that the log-based analysis as an idea/concept had just the ranking value of 5! But each other component of the aggregated value demonstrates that this concept could be realized by the team itself.

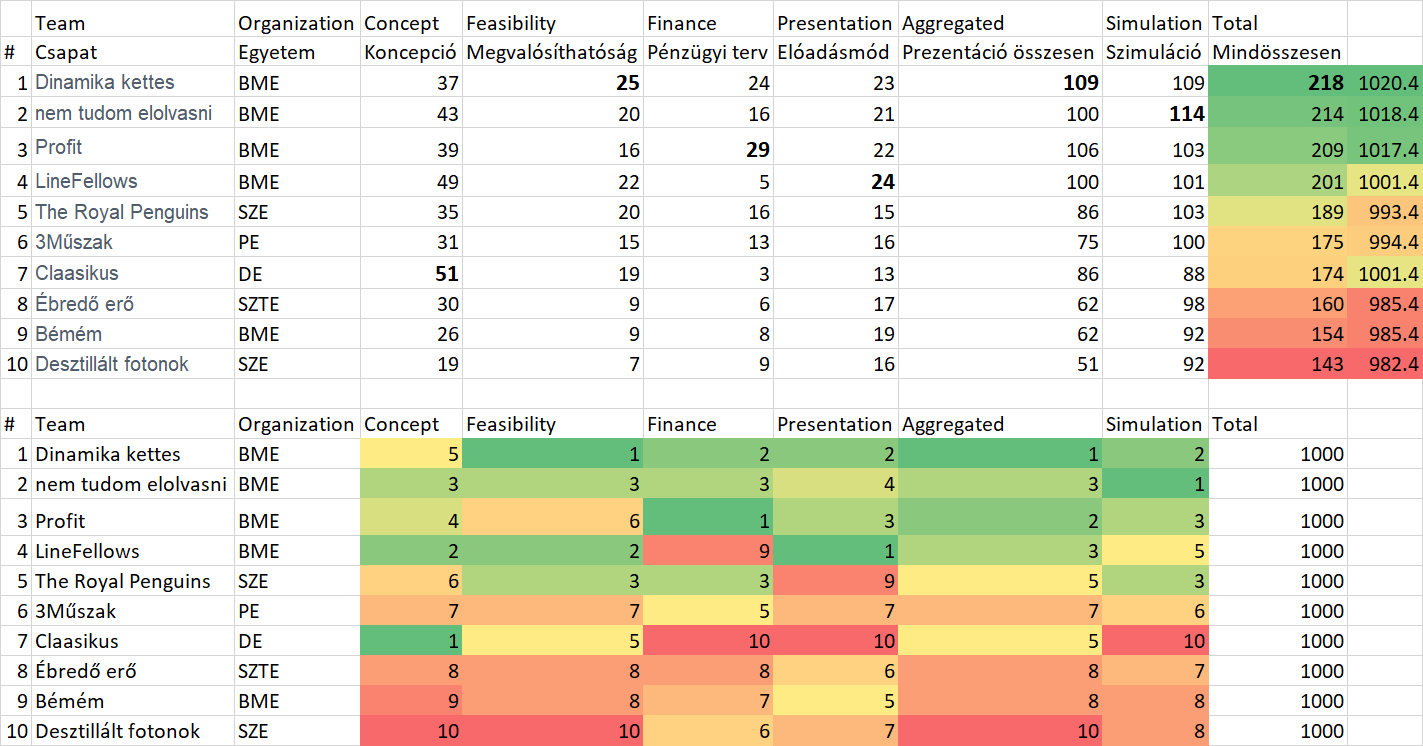


Figure Nr3: Ranking values (source: own presentation - <https://miau.my-x.hu/miau/quilt/UniTech_2019_final.xlsx>)

The Special Award had the team who realized just a ranking value of Nr7. But this team had the best concept. The anti-discriminative similarity analysis (one of the three possible ways for analysing OAMs) could derive the suspicion, that this team should be evaluated with rank Nr4 (parallel to an other team).

So, the evaluation robot (COCO Y0) could be evaluated in a Turing-Test as sensitive enough where the sensitivity means that the robot can see the real values not so like the naïve interpretation of ranking values based on additions and/or averages.

# Steps of the creation of a publication from point of view of a robot-author

The support possibilities of the creation of a publication by Students can be found in the classic form (it means in form of the magic of words):

* <https://miau.my-x.hu/miau/quilt/?C=M;O=D> – demonstration materials: like
  + <https://miau.my-x.hu/miau/quilt/GDP_final_en.doc>
  + <https://miau.my-x.hu/miau/quilt/harmony.docx>
  + <https://miau.my-x.hu/miau/quilt/safety-index.docx>
  + <https://miau.my-x.hu/miau/quilt/20q_based_fingerprints_of_words.docx>
  + <https://miau.my-x.hu/miau/quilt/alternative_evaluations.docx>
* rules/task-interpretations like
  + <https://miau.my-x.hu/mediawiki/index.php/QuILT-IK045-Diary#Synchronicity_Test_Nr.3>
  + <https://miau.my-x.hu/mediawiki/index.php/QuILT-IK057-Diary#Synchronicity_Test_Nr.3>
  + <https://miau.my-x.hu/mediawiki/index.php/QuILT-IK059-Diary#Synchronicity_Test_Nr.3>
  + <https://miau.my-x.hu/mediawiki/index.php/Vita:QuILT-IK045-Diary#8._Day_.282019.IV.03.29>
  + <https://miau.my-x.hu/mediawiki/index.php/Vita:QuILT-IK045-Diary#9._Day_.282019.IV.17.29>
  + <https://miau.my-x.hu/mediawiki/index.php/Vita:QuILT-IK057-Diary#8._Day_.282019.IV.03.29>
  + <https://miau.my-x.hu/mediawiki/index.php/Vita:QuILT-IK057-Diary#9._Day_.282019.IV.17.29>
  + <https://miau.my-x.hu/mediawiki/index.php/Vita:QuILT-IK059-Diary#8._Day_.282019.IV.03.29>
  + <https://miau.my-x.hu/mediawiki/index.php/Vita:QuILT-IK059-Diary#9._Day_.282019.IV.17.29>

The roles and/or the steps are available: <https://miau.my-x.hu/mediawiki/index.php/QuILT-teamwork>

* Thinking experiment
* Searching for data
* Selecting data
* Structuring data
* Deriving reports
* Deriving OAMs
* Modelling
* Interpretation of models
* Creating visual effects
* Publishing

From now on, the paper will present interpretations about the steps above from point of view of a robot-author. **The bold highlighted items are the most preferable option for a robot-author**:

## Thinking experiment

The most complex and yet, most easy subtask is the step of the thinking experiments:

* it is complex because it expects a holistic view at once
* it is easy because a Turing-Test needs just a human-like behaviour pattern not an optimal solution
* therefore, it is possible as a kind of minimum level that a robot
  + identify an arbitrary OAM (even if a keyword like crime is also given)
    - each matrix can be seen as an OAM
    - because an object can be quasi each phenomenon (like time, space, subject, term, etc.)
    - the risk of a randomized selection is minimal because
      * a human being created the matrix (it means this structure should have any trivial interpretability)
      * in case of a matrix having no real meaning, it is possible, that the analytical results bring no relevant declarations what can be seen as a kind of uncovering of the meaninglessness of the OAM
  + choose one of the three analytical ways in a randomized way (like anti-discriminative models, production functions, explorative models)
    - anti-discriminative models can be initialized based on the title of the identified matrix (e.g. service producer prices – time series of countries)
      * <https://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&language=en&pcode=teiis721&plugin=1>
      * <https://ec.europa.eu/eurostat/web/euro-indicators/industry-trade-and-services>
      * [teiis721] - Service producer prices - Index (2015 = 100) - Unadjusted data (i.e. neither seasonally adjusted nor calendar adjusted data) Short Description: The service producer price index (SPPI) is defined as an output price index for the service production of resident producers. It has two main goals. The first is to be a short-term indicator of the business cycle. In this respect SPPI has to reflect changes in prices as fast as possible. The second is to be a suitable deflator for value developments. SPPI should thus represent all output as defined in National Accounts and follow, as closely as possible, the accrual principle in recording prices. Data are not seasonally adjusted.
    - this keyword (service) can be seen as the name of a new index e.g. service-producer-price-index
    - the keyword of time series makes possible to speak about development
    - the keyword of country makes possible to speak about ranking
    - therefore, the title of the study is at once given: e.g.
      * development of the service-producer-price-index
      * ranking of countries concerning their service-producer-price-index-values
    - direction can be declared in different ways: e.g.
      * the correlation-based way:
        + *parallel, the correlations should be calculated between each attribute-pairs*
        + *if the signs of the correlations (of course in the triangle matrix) are always positive at least in case of one of the variables concerning each other variable, then the direction (needed for the initializing of the anti-discriminative model) can be set like the more / the more*
        + *it means: the higher is each attribute’s value the better will be the index as such*
      * the phenomenon-based way:
        + in case e.g. of countries and time periods (see example above) where the number of the classic attributes is just 1 (service-producer-price), the anti-discriminative modelling can speak about

the potential sameness of countries

and/or time-series

* + - * + without needing real directions because the principle the-more-the-more is general valid for time series

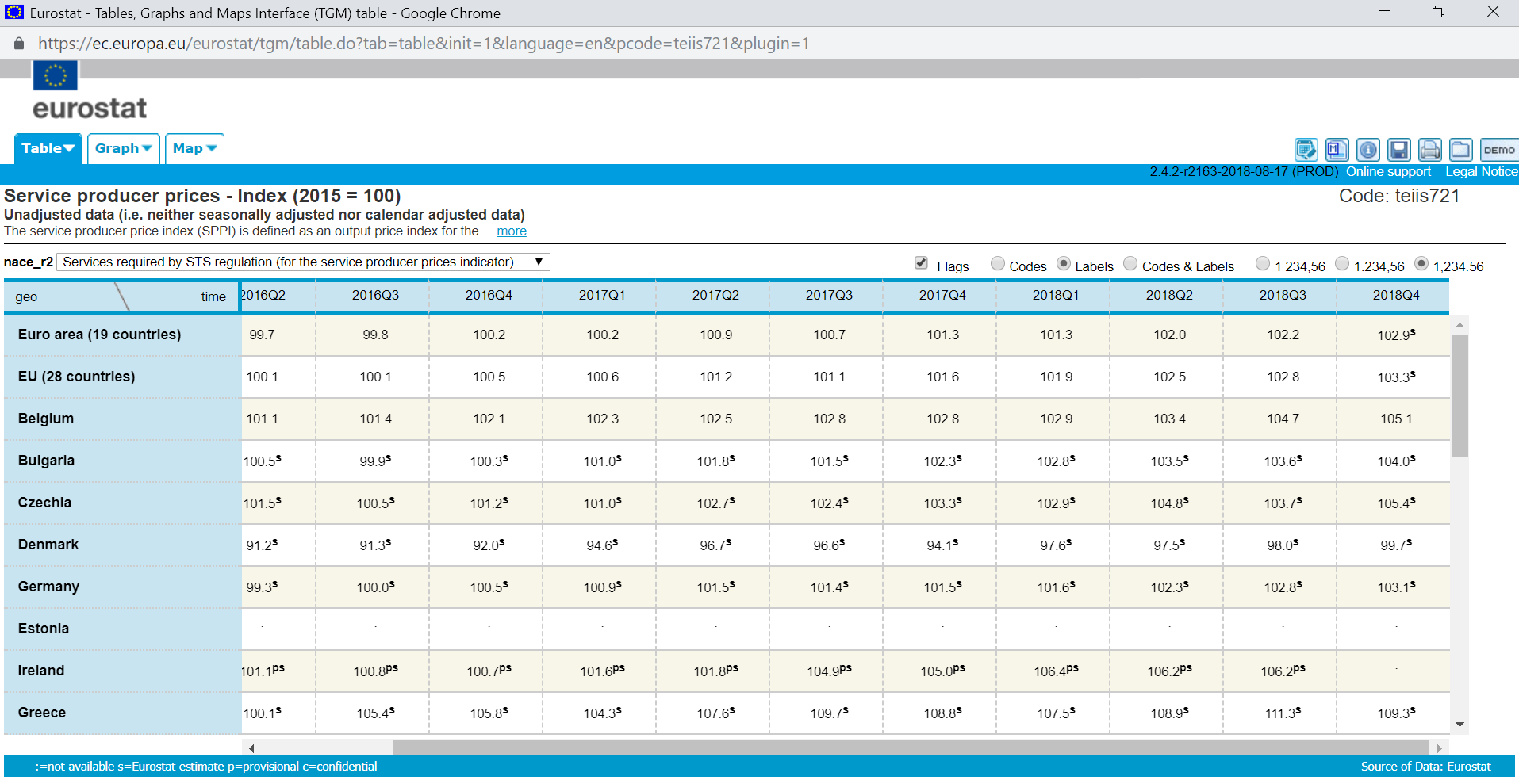


Figure Nr4: Potential OAM (source: EUROSTAT)

* + - production functions can always be interpreted (see Figure Nr4)
      * because one element of the time and/or space can be derived based on the other data
      * it means: the values of one country could be learned in order to estimate one or more lacks of data concerning a country
      * or time periods can also be estimated (forecasted) like countries
      * even both ways could be used parallel to estimate a particular lack in the OAM
    - **explorative models can always be interpreted** (see Figure Nr4 again):
      * the relationships between countries
      * and/or time periods are not trivial…

## Searching for data

As it can be seen, the thinking experiment is not independent from this step: without data, the planned modelling can not be executed. It is possible but mostly on high risk-level, to search for a specific set of data, because the classic statistical services publish just a few OAMs although/because the OLAP-services for deriving flexible reports are available. The parallel way is the interpretation of already finalized OAMs. Its existence is trivial, and it has a further advantage: there is mostly given a previous interpretation which can be discussed. The lack of data (within a matrix) is mostly not a wishful scenario, but it can lead to analytical needs – to estimate the lacking positions.

**Each data set (OAM) can be used for quality assurance analyses where each data position will be interpreted as if the position would not have the appropriate value. This kind of analysis is a part of the explorative modelling. Therefore, in this particular case, the robot-author does not have any problems with contextual layers. The quality assurance models can always be interpreted in a totally context-free way. The quality assurance analysis is important in case of questionnaires where the data can not be accepted as a kind of measured value, therefore, each answer in a questionnaire should be proved whether the answer is a kind of random value (conscious lie, sign of misunderstanding) or the answers and answerers delivered a relative robust value set.**

**Therefore, a low-designed robot-author would prefer the context-free scenarios, which parallel would seem to be especially for human beings as very complex and sophisticated because the human beings believe more frequently on data than it would be acceptable - instead of proving them.**

## Selecting data

**In case of a context-free scenario, it is not relevant, whether an OAM has lacks or not.** In other cases, it is better to have no lacks in the OAM. Robots would prefer one source instead of more sources where the useful combination can not be seen as a context-free scenario.

## Structuring data

**The robot-author would prefer the OAM being identified during the searching. The most trivial identification of an OAM is: to have a HMTL-table, with appropriate tags like <table><tr><td></td></tr></table>.**

Of course, robots can be prepared for other types of matrix-identifications (like XML, JSON, CSV, TXT, etc. – or even Excel-spreadsheets – integrated in PDF or WORD).

## Deriving reports

**Robot-authors do not need reports, if they have the OAM in form of a HTML-table.** But without the appropriate predefined OAM-structure, quasi each 2D-OLAP-report of a database can be interpreted as a rational starting point.

## Deriving OAMs

The reports can mostly be seen as a raw OAM where each attribute can have a different dimension. The real OAM being capable to serve as input for the similarity-based modelling is always a ranked OAM. Ranking values (for each input-attribute) can also be derived online (<https://miau.my-x.hu/myx-free/coco/beker_ranking.php>). Depending on the model-type (Y0, STD, MCM), the OAM should have a specific output variable (Y):

* **in case of Y0: each Y-value should be the same**
* **in case of a standard model: each Y is a real value with appropriate (common) dimension**
* **in case of an explorative model: each Y is a real value**

## Modelling

The modelling itself is just a formal act: the OAM should be transferred to the online engine. This is possible through a HTML-form (<https://miau.my-x.hu/myx-free/coco/index.html>) or through a C-URL-call. The C-URL-calls in the MY-X FREE system are available for partners with big-data-approaches.

The similarity-based modelling is possible in Excel too. The Solver are capable to solve the same problems like e.g. the MY-X FREE services – but Excel is hard limited concerning the volume of the OAMs.

The Google spreadsheets are also able to execute solver-like calls (c.f. NEOS calls), but these services are still not matured enough for real usage.

## Interpretation of models

Parallel to the holistic characteristics of the step about the thinking experiments, the interpretation is a kind of mirrored task of them. The interpretation rules can not be used in form of online engines like the modelling engines. But theoretically, they are also free for everybody: e.g. <https://miau.my-x.hu/myx-free/index.php3?x=t01> The interpretation rules can also be identified in each paper using the particular model-type(s).

## Creating visual effects

The most simple visual support can be realized based on the coloured background of the cells in a spreadsheet-view. The coloured ranking view of an OAM is a specific QR-code being hardly interpretable for human eyes. Robot engines are a kind of robot eyes being capable of seeing pattern where the human eyes can not see them: e.g. <https://miau.my-x.hu/myx-free/index.php3?x=iq>

## Publishing

Expectations (like: <https://miau.my-x.hu/myx-free/index.php3?x=test1>) can be offered for human authors. The same logic can also be used be robots, especially in case of defining text-elements for rules producing papers (c.f. <https://miau.my-x.hu/mediawiki/index.php/Tur_vzsu_tema5_pecs>).

# Conclusions

As it can be seen:

* the expectations for a paper producing a positive Turing-Test seem to be quite simple
  + each matrix can be accepted as a kind of OAM
  + the ranking of the raw values can be executed in form of online services too
  + for each OAM, the modelling approach for quality assurance is a valid challenge
  + modelling engines are automated
  + interpretation rules can be automated
  + useful visualisation effects can be derived in Excel (based on colouring of the background of cells)
  + specific rules for publishing can also be defined and automated…
* each further extension of the expectations needs more and more complex functionalities of the robots (like capability of querying from OLAP-services, detecting HTML-tables, etc.)
* human beings have the massive advantages (even without deep enough education) to be able to handle with the magic of words (it means to react for fuzzy challenges so that it is the benchmark for an appropriate Turing-Test)…