EXPLORING TEXT MINING FOR MEAT RESEARCH TRENDS

Sara Khazzar[1*], Barbara Contiero[1], Ilaria Lanza[1], Giorgia Riuzzi[1],

Stefania Balzan[2], Elisabetta Garbin[1], Lorenzo Serva[1], Enrico Novelli[2]

and Severino Segato[1]

[1]Department of Animal Medicine, Production and Health, University of Padova, Italy

[2]Department of Comparative Biomedicine and Food Science, University of Padova, Italy

*Corresponding author's email: sara.khazzar@phd.unipd.it

I. INTRODUCTION

Among the methods developed for literature data processing, text mining approach helps evaluate big scientific data to find meaningful relationships and trends within specific topics [1]. The study aimed at summarising the outcomes of a text mining and topic modelling analysis of the available scientific literature published over a 15-year period on the three main meat-producing animals (beef cattle, pork and poultry) and meat quality traits.

II. MATERIALS AND METHODS

A literature search protocol was set up to identify abstracts and citations of peer-reviewed papers in the 2006-2020 Scopus® database. Boolean search strings included the following combined key-words: meat, quality, poultry, chicken, swine, pork, beef, cattle, colour, nutritional, organoleptic, shelf life, conventional, organic, intensive, extensive. An inverse document frequency technique (TF-IDF) was applied to weight the number of times a word appeared across all abstracts. A topics modelling analysis was carried out by latent dirichlet allocation (LDA) to generate topic representations according to the frequency distribution of word stems within the investigated papers by an iterative probabilistic process (Gibbs sampling). The topic and statistical analysis were performed by using R package [2].

III. RESULTS AND DISCUSSION

Over the 2006-2020 period, the string used for the literature search identified 1978 records (equal to more than 2 million published peer-reviewed articles). However, 376 records were discharged because they were out of range. Eventually, the 1602 retained records were mainly research articles (88%) and reviews (7%) carried out in a limited number of countries around the world (Fig.1). As expected, the number of publications per year increased over the investigated period (Fig.1).
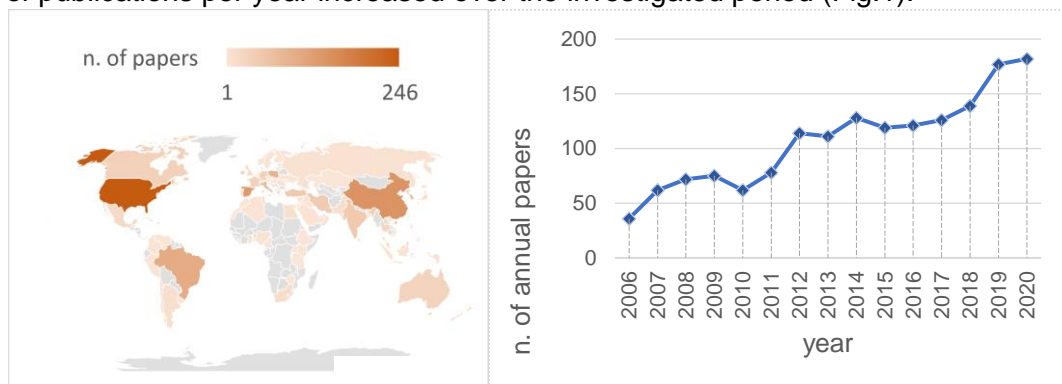

Figure 1. Geographical (left) and yearly (right) distribution of the retained records

According to the TF-IDF ponderation system, the text mining exercise kept 1190 relevant word stems, which are represented as a cloud (Fig.2). The title tentative, the list of the evocative stem terms and the number of retained records of the generated 10 main topics are given in table 1. However, almost

all the topics captured several fields of investigation as many papers focused on multiple-issues making difficult their categorisation within a specific topic. The main research challenges were related to animal and feed performances, while meat safety and shelf life were a second research priority cluster. A large attention was given also to organoleptic quality and meat processing, especially of beef and pork. Over the investigated years, T1 and T2 showed a recent peak in terms of number of published records, while publications within T9 and T3 tended to decline over the last five years (Fig.1).
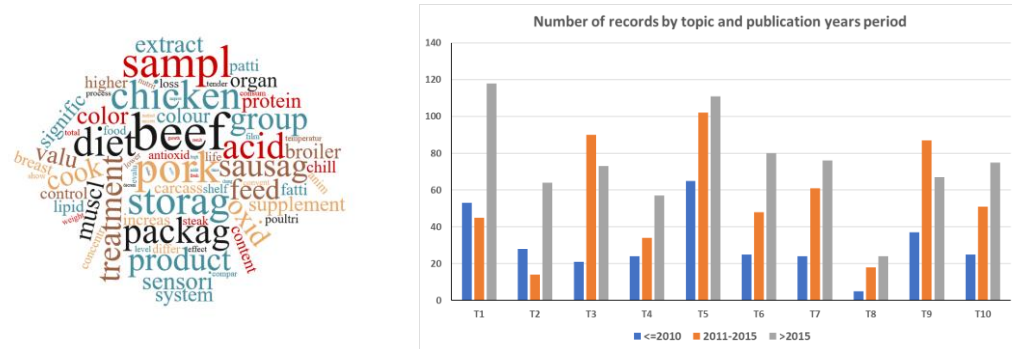

Figure 2. Cloud representation of the most relevant word stems in the database (TF-IDF ≥ 10, on the left); number of records by topic and publication period (on the right)

Table 1 – The most probable word stems defining the 10 topics generated with latent dirichlet allocation

Topic tentative title	Most probable word stems	Records
T1 Meat production	Product food nutrit- poultry use anim- environ- protein consum- can industry import develop- research health include factor main human consid-	216
T2 Assessment & pork	Water chang- use method stud- pork result process model evalu- indic- test analysis applic- can base develop- detect obtain appli-	106
T3 Feed performance	Diet feed supplement fed effect dietary perform level pig treatment increa- organ broiler concentr- protein weight experi- digest miner- growth	184
T4 Beef rearing	Beef differ use cattle anim- fat system carcass trait steer pastur- breed select growth may greater genet- gene associ- produc-	115
T5 Meat safety	Package sampl- day storage shelf sensor- life acid beef microb- count effect color map treatment bacteria store modifi- film	278
T6 Poultry & pig meat	Fatty acid content composit- differ muscl- pig fat slaughter age organ system broiler convent- carcass weight colour rear male bird	153
T7 Organoleptic quality	Loss tender beef muscl- cook temperature effect steak color chill- age increa- time convent- protein shear force display loin heat	161
T8 Performance trial	Signif- chicken group control value show sampl- day compar- total breast differ stud- treatment paramet- observ- result increa- period evalu-	47
T9 Shelf life	Oxid valu- lipid storage extract antioxid- acid active addit- beef effect color oil stabil- pork patti- tbar natur- ground substanc-	191
T10 Processed & sensory	Product sausage sensor- content cook use stud- textur- accept result process attribute property compound characterist- sodium consum- colour	151

IV.   CONCLUSION

The text mining sorted 10 topics related to the research trends for the main meat-producing animals. Although there seemed to be a differentiation among topics, a clear interpretation of the results was made difficult because of a moderate overlapping among word stems. Further explorative text mining should be applied within species and/or the key words combinations should be better refined.

REFERENCES
1. Nalon, E., Contiero, B., Gottardo, F. & Cozzi, G. (2021). The welfare of beef cattle in the scientific literature fron 1990 to 2019: a text mining approach. Frontiers in Veterinary Science 7: article 588749.
2. R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.