Training researchers and research personnel on TDM practice

DOI: 10.4000/books.oep.1752
Publisher: OpenEdition Press
Place of publication: Marseille
Year of publication: 2017
Published on OpenEdition Books: 31 janvier 2017
Serie: Laboratoire d'idées
Electronic ISBN: 9782821878433

http://books.openedition.org

Electronic reference

This text was automatically generated on 16 septembre 2019.
Training researchers and research personnel on TDM practice

1 Researchers must also be trained in the correct use of text and data mining. TDM is currently mostly taught as part of the computer science curriculum, but it is clearly of use to research and its teaching should be extended. In order for all the sciences and every area of research to benefit, it should be taught in as many different curricula as possible.

2 TDM can make the workforce of any country more productive. All the techniques it involves require qualified, or indeed highly qualified, personnel.

Training for the different specialities

3 Concerning infrastructures, TDM will require far more data centres and consequently more technical personnel to develop them: in the construction trades, computing and technical maintenance specialities, and digital engineering.

The emergence of new trades and qualifications

4 Apart from the infrastructures, all the technical trades specific to TDM should also be developed. These professions, specialists in Big Data or database management, are the essential human capital for national economic productivity. The training they receive is therefore crucial.

The initial training of researchers

5 Currently, there are not enough specialised training courses concerning the knowledge resulting from TDM. At university level, the University of Lyon 2 offers a Master’s in Data Mining,¹ and the University of Paris 8 a Master’s in Big Data and Data Mining.² These master’s courses involve training computer engineers competent in “deep learning”, data analysis and management, visual representation of the results of text
mining, etc. In addition to the opportunities of becoming a researcher or lecturer-researcher in these IT disciplines, these training courses also provide entry into the professions of data scientist or study engineer, or statistical study director. The Universities of Paris 6, Nice Sophia-Antipolis and Paris 13 also offer specialised master's courses. However, these courses have a limited number of places (20 in the second-year master's course at the University of Paris 8, for example).

Data-analysis specialities are available as part of certain economics curricula, for instance at the ESSEC (Data Science and Business Analytics) or at the Toulouse Business School (Digital Intelligence and Marketing Intelligence).

However, the demand is increasing and it seems necessary to make these TDM courses available more widely.

Alongside these university or specialised courses, some engineering schools also offer specialist training. For example, the École Polytechnique (in Paris) set up a “data science” department in 2014, in partnership with companies like Keyrus, Orange and Thales. Télécom Paritech is moving in the same direction, as are other schools, but they are distributed too unevenly across France to provide training pathways as rich and popular as in the United States, Germany or the United Kingdom.

Other countries, especially in the English-speaking world, seem to be much more aware of the importance of TDM for the employment market. France has a lot of catching up to do in terms of training, despite the excellence of the existing structures. In Germany there are more master's in computational and data science or in data and knowledge engineering (in Dortmund and Berlin in particular). The United Kingdom and the United States nevertheless have the most advanced infrastructures and training centres in these areas. There are at least twenty different training courses in the United Kingdom, in Coventry, Leicester, London, Edinburgh, Lancaster, Nottingham, Oxford, Bristol, Norwich, Glasgow and Sheffield. Such wide geographical distribution, all over Britain, is echoed in its technical range: theoretical and generalist training (Data Science, Machine Learning), applied training (Applied Data Analytics), and specialised training (Business Intelligence Systems), etc.

In the United States, there are also many training courses in research centres specialising in TDM and Data Science. In particular, this is the case of the University of New York and its Center for Data Science, or the University of Harvard and its Berkman Center for Internet and Society. Several other universities also offer more advanced training in TDM than in France: Stanford, MIT, Northwestern University, Penn State, Indiana, University of Maryland, Carnegie Mellon, Berkeley, etc.

These training courses are essential for better preparing our human capital and also for fostering the increase of knowledge.

**Occupational training and awareness-raising activity**

It is also necessary to train and raise awareness among doctoral students, researchers and lecturer-researchers about depositing and sharing data and about data-processing techniques.

This awareness-raising could be achieved by drafting a good practices guide for Open Science, and by online tutorials and training (e-learning).
This overall training offer will need to be structured and coordinated with the organisations already working in France (Huma-Num, INIST, Persée, CCSD, URFIST, OST, etc.) to provide a national STI training programme for all scientific communities.

NOTES

2. http://www.univ-paris8.fr/Master-MIASHS-Big-Data-et-fouille-de-donnees