



NRC Publications Archive Archives des publications du CNRC

The data age Tapping, Ken

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.

For the publisher's version, please access the DOI link below. / Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

Publisher's version / Version de l'éditeur:

<https://doi.org/10.4224/23000788>

Skygazing: Astronomy through the seasons, 2016-09-20

NRC Publications Record / Notice d'Archives des publications de CNRC:

<https://nrc-publications.canada.ca/eng/view/object/?id=1993a08d-cdf4-4cf1-bf27-561216adb2ae>

<https://publications-cnrc.canada.ca/fra/voir/objet/?id=1993a08d-cdf4-4cf1-bf27-561216adb2ae>

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at

<https://nrc-publications.canada.ca/eng/copyright>

READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site

<https://publications-cnrc.canada.ca/fra/droits>

LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at

PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.



THE DATA AGE

Ken Tapping, 20th September, 2016

A few years ago I was in Ottawa, discussing with an NRC colleague the problems of finding information in large piles of data. Raw data is something like iron ore: not really much use in itself, but the iron we extract from it is very useful. What we try to extract from data is useful information.

The search we were discussing in Ottawa was something like looking in a haystack for things that are not hay, but may look a lot like it. That is, searching for things that do not really fit in with the rest, without knowing beforehand what they are or how they differ. This involves computing techniques such as “data mining”, “neural networks” and “machine learning” – computer aided data analysis (CADA).

In one example, my colleague had taken observations of about 22,000 galaxies and simply asked the computer to divide them into classes, without putting in any information about galaxies or what was being sought. It took a while but the computer did it, and produced a wealth of new astronomical information about those galaxies.

In our project the objective was to find techniques for looking at ice cores, tree rings and sunspot data to put together a very long-term record of the Sun’s behaviour without telling the machine about any particular things we were looking for. In science, as in lots of other human activities, history is filled with examples of missed discoveries because “we knew what we were looking for”.

This is becoming a big issue in astronomy. The amount of data collected in an astronomical observation used to be some notes, a few images, and maybe some spectra. This was usually stored on magnetic tape for us to analyze when we got back home. Modern optical and radio telescopes are not like that. For example the CHIME (Canadian Hydrogen Intensity Mapping Experiment) radio telescope, now well into construction here at DRAO, is capable of imaging a good fraction of the visible sky in a single

operation, observing thousands of objects. The result is that when it starts observing, a torrent of data will be flowing out. The Square Kilometre Array, an international project (including Canada) to build the largest and most sensitive radio telescope ever developed, will produce a tsunami. There will be so much data we won’t be able to store all of it. Somehow we will need initially to process the data in order to reduce the amount without losing scientifically valuable information; in many cases without knowing beforehand what that might be. That’s exactly what CADA does for us.

This problem is not confined to astronomy or even just science. The explosive growth in our capabilities in acquiring all sorts of data, storage capacity, computing power and our ability to move huge amounts of data around the world easily and quickly has led to an accumulation so big the word “huge” is inadequate. Almost anything we want to know is probably “out there somewhere”. The “web browsers” and “search engines” most of us use are tools for finding information that someone has already extracted and processed. Searching the raw data is far more difficult, especially when we have only a hazy idea of what we are looking for. That is where CADA comes in, where the computer becomes more of an assistant than just a number cruncher. Thousands of years ago stone was the technological basis for our ancestors’ lives, so we call that time the Stone Age. Today, living amidst enormous piles of data about everything, including us, we live in the Data Age.

The Sun crosses the equator heading south at 07:21 PST, 09:21 EST on 22 September, marking the Autumn Equinox and the official end of summer. Mars and Saturn lie very low in the southwest after dark. Mars is on the left and Saturn on the right. The Moon will reach Last Quarter on the 23rd.

Ken Tapping is an astronomer with the National Research Council's Dominion Radio Astrophysical Observatory, Penticton, BC, V2A 6J9.

Tel (250) 497-2300, Fax (250) 497-2355

E-mail: ken.tapping@nrc-cnrc.gc.ca

