ATAPY for science

Long-term cooperation between the largest European scientific publisher Springer Verlag and ATAPY Software

Both departments (Software Development & Media Service) of ATAPY Software have been continuously contributing to cooperation with Springer Verlag, the world's second-largest scientific Publishing house, a part of Springer Science+Business Media group which unites 70 publishing companies all over the globe.

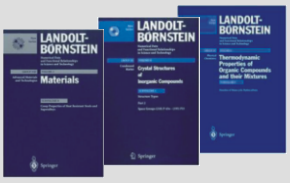
The cooperation started in 2003 with a pilot project intended to assess the efficiency of synergetic use of ABBYY FineReader technology and the expertise of ATAPY engineer-linguists and media operators in digitizing scientific content. ATAPY was entrusted with a small part of Landolt-Börnstein Numerical Data and Functional Relationships in Science and Technology Encyclopedia - a systematic and comprehensive collection of critically assessed data from all fields of physics, physical chemistry, bio- and geophysics, astronomy, materials science, and technology.

The task of converting scientific information into electronic document format is not trivial; prior to contacting ATAPY, Springer had made such attempts, which weren't efficient enough due to out-of-date technology base.

ATAPY successfully accomplished the pilot project which involved TIFF to Microsoft® Word conversion of a series of Encyclopedia pages. The project allowed to understand the nature of the upcoming project, delimit the necessary skills and technologies, understand the main complexities and elaborate solutions for them.

H. Landolt's reference book "Physikalisch-chemische Tabellen" (first issued in 1883 in Germany) presents physicochemical constants of organic and inorganic matters in tabular format. The totally remade 6th edition named "Landolt-Börnstein Zahlenwerte und Funktionen aus Naturwissen-Physik, Chemie, Astronomie, Geophysik und Technik" was issued in 1950-1980.

The appearance of new methods of researches caused the release of "New series" - a reference book named "Landolt Börnstein. New Series. Numerical Data and Functional Relationships in Science and Technology". Since 1961 more than 150 volumes have been issued.



The most serious challenge was a large amount of purely scientific data (formulae, tables, etc.) that contained special symbols missing from the Unicode Character map. This issue was partially overcome by creation of special dictionaries inside ABBYY FineReader - the software package used for full-text recognition - and by implementing a specialized program to help operators promptly insert non-keyboard symbols at verification phase.

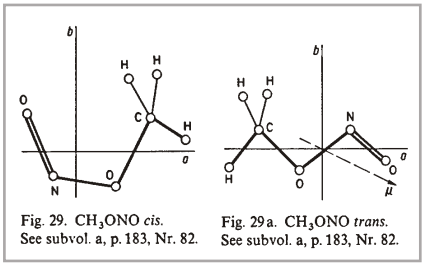


Fig. 29. CH_3ONO *cis*. See subvol. a, p. 183, Nr. 82.

Fig. 29a. CH_3ONO *trans*. See subvol. a, p. 183, Nr. 82.

For analysis of the spectrum the following equation for rotational energy in vibrational state $v_j = v$ is used in most cases

$$W_v(J)/h = B_v[J(J+1)-l^2] - D_v[J(J+1)-l^2]^2 + \Delta . \quad (1)$$

Using $q_j = q_j^{(0)} - q_j^{(1)} J(J+1) + q_j^{(2)}[J(J+1)]^2 - q_j^{(3)}[J(J+1)]^3$,

the following expressions are valid for Δ:

for $v=1$: $\Delta = \pm\frac{q_j}{4}(v+1)J(J+1)$

for $v=2$: $\Delta = \varepsilon\frac{q_j^2 J(J+1)[J(J+1)-2]}{4(B_e - x_{ll})}$

$\varepsilon = 1$ if $l=0$; $\varepsilon = -1$ or 0 if $l = \pm 1$; x_{ll} = anharmonic vibrational constant

for $v>2$: see [A 14].

(1) can be used only if

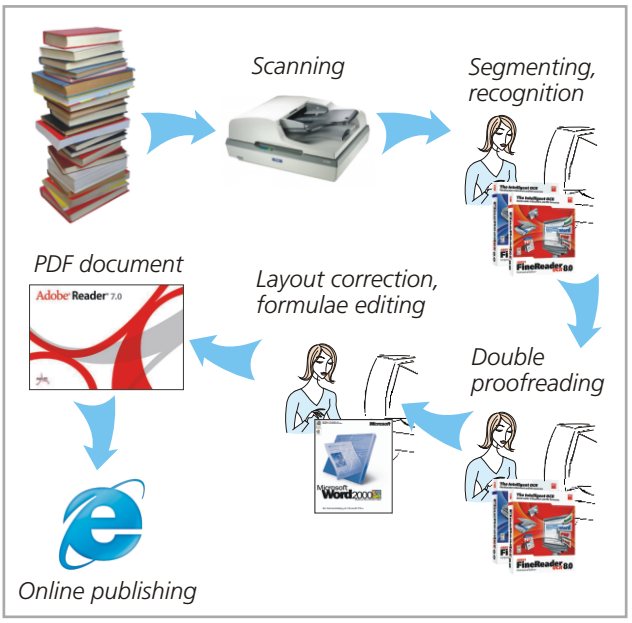
$$B_v = B_e - \sum_j \alpha_j \left(v_j + \frac{d_j}{2}\right) + \sum_{i,j} \gamma_{ij} \left(v_i + \frac{d_i}{2}\right)\left(v_j + \frac{d_j}{2}\right) + \gamma_{ll} l^2 \quad (2)$$

The second goal to achieve was an accuracy level of over 99.99%, which meant less than one mistake per 10,000 characters. The goal was achieved with the use of excellent OCR technologies of ABBYY Software House adjusted for this particular task by ATAPY, and due to the meticulous verification work of ATAPY Media Service department.

This promising beginning grew into full-fledged cooperation between Springer and ATAPY.

The second project involved digitization of larger amounts of the same Edition with further conversion into A++, the customer-specific XML format. As a part of this task, ATAPY engineers automated the A++ conversion of the reference lists following each chapter of the Edition.

ATAPY Software also worked on digitization of the numerous charts „populating" the Edition in order to allow the material's online usage and eventual interactivity for scientists of the XXI century.



Springer Science+Business Media, or Springer, is a worldwide Publishing house based in Germany which publishes textbooks, academic reference books, and peer-reviewed topical journals with a focus on science, technology, mathematics, and medicine. Within the science, technology, and medicine sector, Springer is the largest book publisher, and second-largest journal publisher worldwide, with over 60 publishing houses, 1,900 journals, 5,500 new books published each year, sales of 924 million euro (in 2006) and 5,000 employees. Springer has major offices in Berlin, Heidelberg, Dordrecht (the Netherlands) and New York.

ATAPY Software

630090, Enginernaya Street, 4a, 522
Novosibirsk, Russia
Tel. +7 383 33 56 569 Fax +7 383 33 56 561
www.atapy.com office@atapy.com

Springer Science+Business Media

Heidelberger Platz 3 14197
Berlin Deutschland
Tel. +49 6221 487 0 Fax +49 6221 345 0
www.springer.com