Computers and Chemical Engineering 35 (2011) 671–678

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/compchemeng

Data flow modeling, data mining and QSAR in high-throughput discovery of functional nanomaterials

Yang Yang[a], Tian Lin[a], Xiao L. Weng[b], Jawwad A. Darr[b], Xue Z. Wang[a,*]

[a] Institute of Particle Science and Engineering, School of Process, Environmental and Materials Engineering, University of Leeds, Leeds LS2 9JT, UK
[b] Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, UK

ARTICLE INFO

Article history:
Received 4 December 2009
Received in revised form 25 March 2010
Accepted 22 April 2010
Available online 26 May 2010

Keywords:
Data mining
QSAR
Design of experiments
Genetic algorithm
Nanoparticle
High-throughput

ABSTRACT

Metal oxide nanoparticles are promising materials in applications for fuel cells, gas sensors and fine chemical catalysis. Their functionality depends excessively on composition, structure as well as synthesis and processing conditions. Continuous hydrothermal flow synthesis (CHFS) reactors are an effective technology to make nanoceramics. In order to increase sample throughput of CHFS, a manual high-throughput continuous hydrothermal (HiTCH) flow synthesis process capable of formulating scores of samples per day was developed. More recently, a fully automated nanoceramics synthesis platform called RAMSI (rapid automated synthesis instrument) based on the HiTCH synthesis technology was developed. When large numbers of nanoceramics are made and formulated into appropriate libraries, automated analytical instruments can be used to allow collection of a large amount of useful data. This paper describes the information flow management system of RAMSI (as well as CHFS) and the data mining system for supporting discovery, QSAR (quantitative structure–activity relationship) modeling and DoE (design of experiments). Case studies demonstrating the use of the high-throughput data mining system are presented. These include clustering of Raman spectra, interpretation of X-ray diffraction (XRD) measurements, and QSAR model building linking XRD data and photocatalytic properties. A genetic algorithm method for DoE is also presented that can guide the experiments to search optimal XRD patterns.

1. Introduction

Nanoceramic particles are prospective materials in applications for solar cells (Oregan & Gratzel, 1991), fuel cells (Brandon, Skinner, & Steele, 2003), gas sensors (Ge, Xie, & Cai, 2007) and fine chemical catalysis (Kaspar & Fornasiero, 2003). The performance of new functional nanomaterials is very sensitive to dopant cations levels (Zhang, Brown, et al., 2009). In particular, doped nanomaterials can show large changes in properties due to small differences in particle size or composition (Bognolo, 2003; Hyeon, 2003). Due to a lack of theoretical and priori knowledge, the understanding on how nanoparticle properties vary with composition and structure is limited. The use of experiments to search for new nanomaterials is severely hindered by the speed at which new compositions can be made and tested for suitable properties (Hagfeldt & Gratzel, 1995).

Some of the authors developed continuous hydrothermal flow synthesis (CHFS) reactors (Boldrin et al., 2007; Chaudhry et al., 2008; Chaudhry et al., 2006; Thompson et al., 2009; Weng, Boldrin, Abrahams, Skinner, & Darr, 2007; Weng et al., 2008; Weng, Perston, et al., 2009; Zhang, Brown, et al., 2009; Zhang, Goodall, et al., 2009; Zhang et al., 2009) and manual high-throughput CHFS known as HiTCH (high-throughput continuous hydrothermal) which were used to make 66 samples in a phase diagram (Weng, Cockcroft, et al., 2009). The basic HiTCH process (Weng, Cockcroft, et al., 2009) is shown in Fig. 1. A supercritical water feed (ca. 450 °C and ca. 24 MPa) is pumped towards a reaction point (R), where the water can meet with a flowing mixture of cold pre-mixed metal salts (delivered via a syringe pump) and the auxiliary feed such as base. At the reaction point (R), nanoparticle slurries are formed and the slurry containing nanoparticles are then cooled after passing through a heat exchanger (T), and collected after the back-pressure regulator (B) into tubes. The system is then purged before a new mixture of metal salts is added (samples may also be sent through the system as a purge prior to collection of products also). Any collected nanoparticle samples are then cleaned by replacing the supernatant with clean water several times. Samples can then be printed onto hydrophobic paper as spots which dry to form compact ceramic dots if required.

After drying and firing under different temperatures if needed, the dots can be embedded into wellplates as combinatorial libraries (Lin et al., 2010) for batched analysis of properties, such as X-ray diffraction (XRD), Raman spectroscopy, luminescence spec-

* Corresponding author. Tel.: +44 113 343 2427; fax: +44 113 343 2405.
E-mail address: x.z.wang@leeds.ac.uk (X.Z. Wang).

0098-1354/$ – see front matter
doi:10.1016/j.compchemeng.2010.04.018

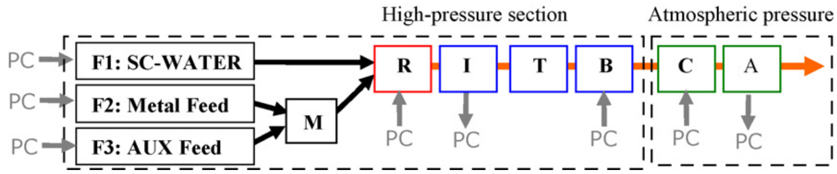

Fig. 1. Scheme of CHFS or HiTCH synthesis for nanoparticles. R, reaction point; M, mixing T-piece; I, in-line monitoring of temperature and pressure; T, heat exchanger for cooling; B, back-pressure regulator; C, sample collection, cleanup and printing; A, batched analyses and photocatalytic tests (XRD, Raman, UV/vis, BET, and dye decolourisation, etc.).

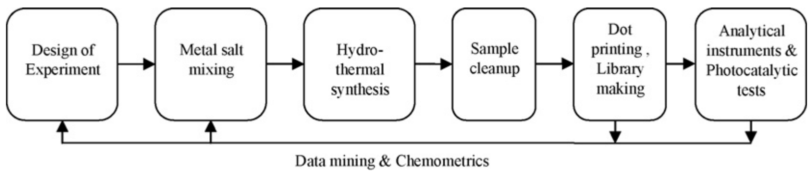

Fig. 2. Work flow of high-throughput nanomaterials discovery.

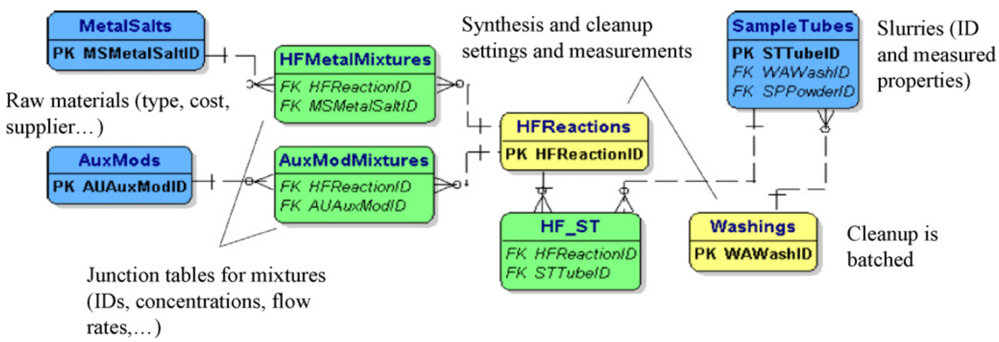

Fig. 3. Data structure design for synthesis section showing entities, key attributes and relationships between entities. Yellow tables are the HT process (drying, clean up, reaction,. . .); the blue tables present all the physic things (tubes, materials, equipments,. . .); the green ones are junction tables. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

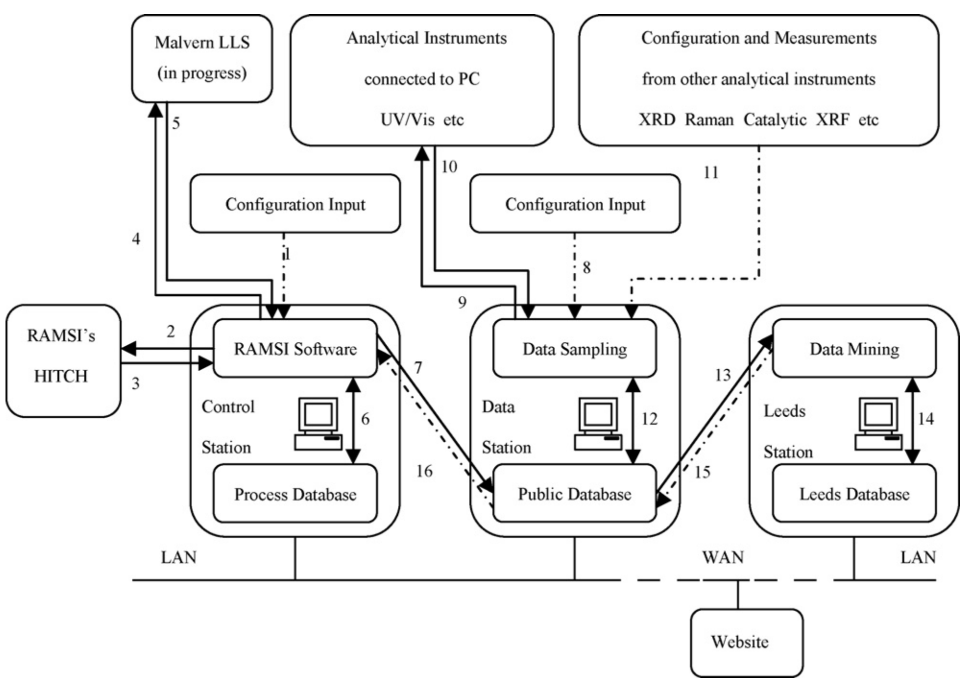

Fig. 4. Process database controlled by RAMSI software (flows 1–6). Synthesis and storage data copied into University College's (UCL's) database (flow 7). Analytical data added by PCs running instruments (flows 10, 11). Updated data of UCL database will be captured by Leeds database and data mining support and feedback information will be sent to UCL station through InforSense (flow 13, 14, 15). Revisional experimental design information will be sent to the RAMSI control software (flow 16).