

Curriculum vitae Pieter Masereeuw

General data

Name Masereeuw, Pieter Christiaan
Date of birth November 5, 1957 (Amsterdam)
Civil status Married, 4 children
Address Anna van Saksenstraat 25; 1901 TH Castricum
Phone/mobile +31 251 670964, +31 6 20786265
Email pieter@masereeuw.nl
See also <http://www.masereeuw.nl> (in Dutch)
Picture



Education and experience

High school 1970-1976. Grammar-school
(French, German, English,
Greek, Latin, Dutch, History)

B.A. University of Amsterdam,
1976-1981, cum laude (classical
languages)

M.A. University of Amsterdam,
1981-1988, cum laude (Latin
linguistics, Computer Science
and General Linguistics)

Programming and scripting
languages Java, Javascript, C++, C, Pascal/
Delphi, Prolog, Smalltalk, Perl,
Unix shells, JSP, PHP, ACL
(Arbortext Command Language),
8086-Assembler

Platforms Unix (Solaris, Ultrix, AIX, Linux),
VMS, AOS/VS, Macintosh,
Windows NT

Formalisms SGML, HTML, XML, XSLT, SVG,
XSL-FO, XML Schema, lex/yacc,
javaCC, SQL, ..

Overview of current and past jobs

1/1/2008-	Freelance work via <i>ZZP Oké Pieter Masereeuw</i> . A list of customers: Sanoma Learning, Uitgeverij Malmberg, Ten Brink Offset, NEN (standardization), Ambrac/Boom Uitgevers Den Haag, Ambrac/Kluwer, D-reizen, Uitgeverij Pegasus, Politieacademie, Nationale Politie, DCT (a.o.. for Vanderlande Industries, DAF, Nedtrain and Damen Shipyards), CAK (via Informaat), Instituut voor Nederlandse Lexicologie, iCtrl (for Lloyds Register Rail Europe and for Nedtrain), Uitgeverij Springer/BSL, The Docworkers, SDU, Van Dale Lexicografie, Uitgeverij Thieme, Logica (for ING/Postbank), TNT Post Electronisch zakendoen,
1/1/2004-31/1/2008	ICT coordinator for Alfabase in Alphen aan den Rijn. Also 1 day per week freelance work.
1/11/2000-31-/12/2003	IT Team leader at PlantijnCasparie Data in Heerhugowaard (4 days a week). Also 1 day per week freelance work. Job ended because of company close down.
1/06/1999 - 31/10/2000	Software engineer at Ambrac B.V. in Utrecht.
1/01/1997 - 30/05/1999	System analyst/programmer at the Institute for Dutch Lexicology (INL) in Leiden.
1/12/1995 - 31/12/1996	System developer at Informaat in Baarn.
1/11/1986 - 1/12/1995	System analyst/programmer at the Computer department of the Arts Faculty of the University of Amsterdam.
1/03/1984 - 1/11/1986	Technical programmer at two research project founded by the Dutch Government (ZWO): Conversion of the Longman Dictionary of Contemporary English and its application to automatic parsing of English language texts.

1/05/1983 - 1/07/1985	Programmer and system manager at the Computer department of the Arts Faculty of the University of Amsterdam.
1/09/1979 - 1/09/1983	Research assistant at the Latin department of the University of Amsterdam, with the task to develop computer software for the retrieval of linguistic phenomena from Latin text corpora.
1/09/1979 - 1/03/1984	Student job: PTT Telephone Exchange in Amsterdam (international telephone connections).

Self-employed via *ZZP Oké Pieter Masereeuw*, starting at January 2008

As a freelance worker, I aim at obtaining assignments that match my experience and interest areas. These areas are mainly text technology but sometimes also language technology. Examples of these are:

- Projects pertaining to the development of software for editorial activities, such as data entry, conversion, and DTD/Schema-design)
- Information retrieval from texts (free-text search, dictionaries)
- Conversion to publication formats, for print as well as electronic publications. In the case of conversion to print, I like to co-operate with people that specialize in InDesign, FrameMaker, XyVision and 3B2. Conversions by means of XSL-FO are done by myself.

For this kind of projects, I prefer to use open standards, among which XML, SGML, XSLT, XSL-FO, SVG and XHTML. For the development of computer programs I prefer development environments that allow my software, including a GUI, to run on various platforms. This implies that, apart from webbased interfaces and XSLT, Java/Java Swing is my favourite programming platform.

My most recent employers:

- **KOOP (Knowledge centre and Exploitation Center for Official Dutch Government Publications).** System manager and developer for various government websites and retrieval systems for government publications. Techniques: XML, Schema, XSLT, XSLWeb, Solr Cloud, Interactive XSLT, JSON, Apache configuration, webservices (REST), low level HTTP.

- **Uitgeverij Malmberg** - programming an XSLT conversion for student courses about math; also development of XML Schemata, Schematron and XQuery scripts. Conversion of Word and other Office documents to Malmberg/Sanoma's proprietary XML format. Development of Java programs for an intelligent ZIP tool and the management of persistent identifiers.
- **Van Gogh Museum** - conversion of Word documents to TEI XML for the diaries of Jo Bonger, Theo van Gogh's wife. Additionally, I developed some editorial tools for work in Oxygen XML Editor.
- **International Baccalaureat (IBO)** - development of an XSLT tool to convert Word files to JATS XML, plus the creation of a web application in order to manage conversions.
- **EZ-base** - modification of XSLT-stylesheets that render XML data into PDF catalogue files by means of XSL-FO.
- **Instituut voor de Nederlandse Taal (INT)** - development of a Rich Internet Application for the historical dictionaries of the institute, based on Interactive XSLT (Saxon-JS) and Bootstrap. See <http://gtb.ivdnt.org>. Also, many conversion scripts based on XSLT and XProc for the conversion of files to TEI format.
- **Juridisch Woordenboek Spaans** (privately funded) - design of the dictionary schema (XSD), creating a plugin for Oxygen XML Editor (Oxygen Java API, XSLT, CSS)) for editors who are not really computer-savvy, set-up ExistDB in order to automatically generate reports about the state of affairs of the editorial work flow..
- **Fryske Akademy (Frisian Academy)** - setting up/ adapting an editorial environment for a Dutch-Frisian dictionary (Java, MySQL, Schema, XSLT) and publishing the dictionary by means of a web interface.
- **Sanoma Learning** - conversion of XML files to JSON format for use in Sanoma courseware.
- **DCT** - Converting a large set of office documents (Word, Excel) to DITA in a fully automated fashion and then publishing the generated set of DITA files to large PDF documents by means of a customized DITA Open Toolkit.
- **NEN** - creating an XSLT stylesheet in order to render NEN documents as HTML.
- **Ten Brink Offset** - assistance in the development of various order intake systems and workflow systems, in Java.
- **Ambrac/Boom Uitgevers Den Haag** - creation of a Solr indexing system using Jenkins, rsync and Linux shell scripts.
- **Ambrac/Kluwer Law International** - reverse engineering the scripts that are used inside the Sigmalink Content Management System.

- **Pegasus Uitgeverij** - technical XML-support for various dictionaries.
- **Police academy** - creating a DITA (XML) specialization for the Police Academy. Configuring an XML editor (XMetaL, Oxygen), converting existing text material to the specialized DITA format. Participate in the selection process for a content management system.
- **DCT** - Proof of concept: publishing Dita Topics via Sharepoint and Dita Exchange for a national energy distribution company.
- **CAK** (via Informaat and BeInformed): XSLT-work (including CSS and some Javascript) for the CAK's BeInformed web portal and the generation of standard letters and other documents (XSL-FO).
- **DCT** - Development of a Javascript application for Damen Shipyards that enables navigation (!) on sea in the set of manuals that go with the ship's systems.
- **DCT** - Adapting the DITA Open Toolkit (Webhelp and PDF via XSL-FO) for Damen Shipyards.
- **DCT** - Making a DITA specialization for Nedtrain and adapting the DITA Open Toolkit.
- **DCT** - Modification of complex XSL-FO tables for Vanderlande Industries.
- **INL (Institute for Dutch Lexicology):**
 - Advice on the use of DITA and DITA editors for a large linguistic project
 - Conversion of many formats (including newspapers) to the TEI DTD
 - Creation of an interactive correction tool for TEI documents, using a Java Swing interface
- **iCtrl** - building a a component based XML editorial system for the Dutch Train maintenance manuals of Nedtrain. The system uses XMLMind (XML editor), Ovidius TCToolbox (CMS), XSLT 2.0, CSS, Java and Prince XML (for PDF rendering). The online version consists of an HTML interface that uses Javascript/JQuery that helps comparing different versions of the document.
- **iCtrl** - development of an XML Schema and configuration of an XML Editor (XMLMind) using Java. Purpose of the project is cleaning up a database of a supplier of NS (Nederlandse Spoorwegen, Dutch Railways).
- **Springer/BSL (Bohn Stafleu van Loghum)** - XML conversion to the ePub-format for e-readers - the creation of a tool that performs the conversion, and is equipped with a Java Swing user interface.
- **Van Dale Lexicografie**
 - Conversion of Oxford Dictionaries to publication format (XSLT, CSS)
 - technical/editorial advice;
 - schema development; conversion of various internal formats (using XSLT 2.0) to one XML-schema (W3C XSD);

- Interactive tools for the exploration of large dictionary files, using SAX and Java Swing technology
- **The Docworkers**
 - conversion of 3B2 files to XML;
 - development of an XML roundtrip (Word-DocBook-Word).
 - Software for creating e-books in the ePub format, with a Swing-based user interface
- **SDU** - DTDs and upconversion for Dutch electronic legislative documents;
- **Thieme** - solving a bug in an Apache Cocoon pipeline (fixing the XSLT component);
- **Logica (for ING/Postbank)** - software fixes for the integration of the software of ING and Postbank (C++, Java, Easel);
- **TNT Post Electronisch Zakendoen** - automatically adapting a demo website for electronic invoicing (Java, JBoss, Seam), based on a Java Webstart Swing-application;
- **Sheridan Europe** - Java programming for international job distribution for Printing on Demand (POD) using low level Java APIs such as mail, ftp, pdf-manipulation/itext and Java Swing for Webstart GUIs.

Alfabase Cross Media Solutions, 2004-January 2008

Alfabase Cross Media Solutions is part of the American-European Sheridan group. The company distinguishes itself from ordinary printing and POD companies by advanced prepublishing activities and a high degree of automatization. Other automatization activities are automatic order intake and order processing.

At Alfabase Cross Media Solutions, prepublishing activities are focused on the automatical (without human intervention) formatting of structured data (XML, SGML) into attractively rendered PDF files. Another activity is making the manual publishing process (DTP) smarter by means of scripting and tools for co-operation via the internet.

Software for automatic order intake and order processing is created mainly for POD (printing on demand) activities of other companies within the Sheridan group. It is an international activity that enables customers (large publishing houses) to send their files to a central place, after which print jobs are distributed over the various Sheridan companies in the world.

This leads to a large reduction of distribution times and expenses.

My task at Alfabase and Sheridan was the co-ordination of ICT development activities. These activities comprised:

- Carrying out research regarding prepublishing, automatic manipulation of PDF files and order processing.
- Managing a group of software developers.
- Contribute, as a member of the management team, to the direction that Alfabase and Sheridan should take with respect to prepublishing and workflow automatizations. On both areas, I closely work together with our German sister company [CrossMediaSolutions](#).
- Conducting software design and development; an important product was xzPages, which combined a Swing-based Java Webstart client with Unix server scripts for automated typesetting.
- Supporting commercial activities, for instance giving demos and explaining the benefits of our approach to the customer, especially concerning editorial activities.

Freelance activities, one day a week, during this period (2004-January 2008)

My freelance activities included:

- Teaching about XML and related standards.
- Internet applications (JSP) for the registration of the availability of part time teachers of several primary schools in my region.
- Porting of an application (C, Pascal, lex, yacc to JavaCC and Java) for the phonological analysis of the speech of language-impaired children. This application will be based on Java Webstart technology and will be made available to clinical speech trainers. The original version of this software, called FAN, was developed by me at the University of Amsterdam, but I obtained the rights. See also [below](#).
- Programming activities in Delphi, C, C++ and Java. One of the jobs was porting the University of Amsterdam parser generator [Atlas](#) from Unix to Windows, and supplying it with a Java Swing user interface.
- Various small internet applications, such as an application to keep the scores of the North-Holland-North Darts organization (in PHP) and a Swing-tool for randomly selection opponents in the competition.

Other job-related activities

- Editor of <!ELEMENT, the magazine of the SGML/XML Users Group Holland).
- Member of the SIG (Special Interest Group) about XSL-FO of the above-mentioned Users Group.
- Participate in the organization of a workshop about up conversion to XML by means of format grammars, given at the SGML/XML congress 2006.
- Organization of a workshop about XSL-FO on the technical day of the SGML/XML-congres 2004.
- Organization of a workshop about [Apache Cocoon](#) on the technical day at the SGML/XML congress 2003.

PlantijnCasparie Heerhugowaard, 2000-2003

PlantijnCasparie Heerhugowaard was a company that had more or less the same activities as my later employer Alfabase. Apart from being a printing company, it specialized in prepublishing and automatization. Unfortunately, the owner of the company made an end to this activities and closed down the Heerhugowaard company. I left before the closing down became effective.

At PlantijnCasparie I was team leader of a team of software developers and a web designer. Apart from that, I was involved in programming activities.

I was also hired to conversion projects at Kluwer Alphen/Deventer and to a project where I created the interface logic for a web application of a route planner that took congestion expectancies into account (at Rijkswaterstaat, part of the Dutch government).

Activities

At PlantijnCasparie my team and I were involved in the following activities:

- *E-commerce*
The internet applications were initially developed in JSP pages with XML, but later we made a switch to Apache Cocoon.
- *Dictionaries and encyclopedias and CD-ROM and internet*
An example is the CD-ROM that goes with the Latin-Dutch dictionary (the website <http://www.latijnnederlands.nl> is based on the design of this CD-ROM). For the creation of internet applications,

we used Java Server Pages (JSP) and many other Apache projects: Apache Tomcat, Apache Cocoon, Apache FOP and Apache Lucene.

- *SGML/XML applications*
On the one hand, our software supported editorial activities and on the other hand it enabled the automatization of the typesetting process. Automated typesetting was done using a typesetting system (3B2), but also (experimentally) with XSL-FO.
- *Order intake*
Adapting an interactive (Java Swing) program used by order managers to track the production of SGML- and PDF-processing of scientific journals.

Ambrac, 1999-2000

The Ambrac company renders services for the processing of XML and SGML. At Ambrac, I was involved in the following activities:

- *Content management*
Configuring the *Sigmalink* content management system for Kluwer Law International in The Hague.
- *Conversions*
On the one hand I developed Java-based regular expression tools that made the creation of conversion tools easier. On the other hand, I used standard tools such as Perl and XSLT (then being in a draft version).
- *Editorial tools*
The creation of scripts in *Adept*Editor*, mainly in order to create a flexible editorial environment at Samsom publishers (nowadays called Kluwer Alphen).

Institute for Dutch Lexicology (INL), 1997-1999

At this institute, I was the single developer. As such, I was involved in:

- *Corpus retrieval software*
At the INL, I developed new corpus retrieval software, financed by the European Union. Some properties of this program were:
 - Client/server: the client in Java (applets), the server in Java and C++;
 - Object orientation; this created a large degree of configurability. Even the query language was a parameter;

- Java- applets (for the client);
 - Servlets (then being state of the art; JSP and other Java techniques were not yet there);
 - Distributed applications: via sockets, the server commands processes on various other machines;
 - Multithreading, in C++ (Posix-threads) and in Java;
 - SGML/XML-aware searching techniques;
 - > 40 miljoen words;
 - A combination of searching by means of bit vectors and regular expressions.
- *Systems for conversion and validation*
Especially for the creation of conversion programs, I developed my own scripting language (*Taggle*). The most obvious tool - *Perl* - proved to be less adequate - many operations needed to operate within certain limited subsections of the text, which is rather involved in Perl. The Taggle languages also proved a handy tool for HTML-conversion en CGI-scripts. Taggle was written in Gnu Bison (yacc), flex en Gnu C. Versions were created for Sun/Solaris, Linux and (experimentally) Windows NT. A Taggle script was compiled to a C program, so ultimately, it produced native code.
 - *Lexicographical software*
New versions of the so-called *Groene boekje* (the official word list for Dutch orthography) were mainly created by means of Taggle scripts.

Informaat, 1995-1996

At Informaat, I participated in the development of a system (*Dox*) that allowed company information to be entered in a structured way, in order to exported to various media, such as print, RTF-files and hypertext files.

Development was carried out in the mother of all object oriented languages, *Smalltalk*.

University of Amsterdam: Department of Latin and Computer Department of the Faculty of Arts, 1979-1995

Computer Department of the Faculty of Arts rendered services to the Arts faculty and the university as a whole. At the time when computers were not yet seen anywhere, it offered computing facilities and standard software (often home-made) for research in the faculty. I was involved in the following projects:

- *Retrieval from large text corpora*
By means of this program, developed by G.J. van der Steen, linguists and literature scientists could test their hypotheses about language and literary phenomena in corpora - large unstructured bodies of natural language samples. In order to make searching more effective, we developed programs to apply linguistic tagging to the words.
- *Systems for conversion and validation*
Tothether with a team of programmers, I was involved in the development of a parser generator for formal grammars: *Atlas* (earlier: *Parspat*, and nowadays *AddXML*). This parsing tool was not only used for analysis and input checking, but it was also able to perform rule based conversion (aka *transduction*). My thesis for Latin linguistics was such an application: it is a context sensitive computer grammar for Latin morphology which is able to automatically stem and tag Latin words. This product, called *Latinflexion* was sold to the Katholische Universität Eichstätt (Germany) and to the department of Latin of the University of Amsterdam, together with an interactive program that supported manual disambiguation of the output.
- *Well-structured data entry*
The development of *DictEdit*, a Macintosh program for the consistent entry of dictionaries, was my most import activity in the period from 1992 to 1995. Based on a definition file, dictionary input forms were presented in a Macintosh window. The data that was entered was stored in SGML format. By means of a transformation engine, print-like Dictionary entries could be created in another window. DictEdit was presented during the Euralex congress about lexicology at Tampere (Finland).
- *Fan*
Phonological/logopedical analysis of Dutch utterances by language learning young children. Upon leaving the University, I bought the rights of this program which is now, on a limited scale, being distributed my me and my former colleague. Currently, a new Java Webstart based version is being developed.

- *CM/LTree*
For many applications in the time when computer memory was expensive, this library was very beneficiary. CM is a portable cache-memory library which I wrote in Ansi C, and LTree (which builds upon CM) is a C library for the storage, indexing and retrieval of large lexical databases. It has been used by many internal projects, and also by EU financed dictionary projects, such as *Acquilex* and *Sift*. The software runs under Unix/Linux, Windows and Macintosh.

Other work-related activities

- Member of a workgroup of several Dutch universities for the development of software for computing in the Humanities.
- Member of the board/treasurer of the STDH ("Dutch Society for Text Corpora and Data files in the Humanities").