

ISBN 978-0-626-32084-3

SANS 24614-1:2015

Edition 1

ISO 24614-1:2010

Edition 1

SOUTH AFRICAN NATIONAL STANDARD

Language resource management — Word segmentation of written texts

Part 1: Basic concepts and general principles

This national standard is the identical implementation of ISO 24614-1:2010, and is adopted with the permission of the International Organization for Standardization.

SANS 24614-1:2015

Edition 1

ISO 24614-1:2010

Edition 1

Table of changes

Change No.	Date	Scope

National foreword

This South African standard was approved by National Committee SABS/TC 037, *Terminology and other language and content resources*, in accordance with procedures of the SABS Standards Division, in compliance with annex 3 of the WTO/TBT agreement.

This document was approved for publication in September 2015.

SANS 24614-1:2015
**INTERNATIONAL
STANDARD**

**ISO
24614-1**

First edition
2010-11-01

**Language resource management — Word
segmentation of written texts —**

Part 1:

Basic concepts and general principles

*Gestion des ressources langagières — Segmentation des mots dans
les textes écrits —*

Partie 1: Notions fondamentales et principes généraux



Reference number
ISO 24614-1:2010(E)

© ISO 2010

PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.



COPYRIGHT PROTECTED DOCUMENT

© ISO 2010

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword	iv
Introduction	v
1 Scope	1
2 Terms and definitions	2
3 Basic framework for word segmentation	6
4 General principles of word segmentation	10
Annex A (informative) Representing word segmentation in XML	13
Bibliography	14