

Perceptive Search

Release Notes

ImageNow Version: 10.2.x

Written by: Product Documentation, R&D
Date: September 2016

perceptivesoftware
from Lexmark

© 2014 Perceptive Software. All rights reserved.

Perceptive Software is a trademark of Lexmark International Technology S.A., registered in the U.S. and other countries. All other brands and product names mentioned in this document are trademarks or registered trademarks of their respective owners. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or any other media embodiments now known or hereafter to become known without the prior written permission of Lexmark.

Table of Contents

Perceptive Search 10.2 Release Notes (January 29, 2013)	4
--	----------

Perceptive Search 10.2 Release Notes (January 29, 2013)

- Perceptive Enterprise Search (formerly ISYS Enterprise Server) 10.2
- Perceptive Workgroup Search (formerly ISYS Workgroup) 10.2

If you are upgrading from a previous version of ISYS Enterprise Server or ISYS Workgroup, refer to the *Perceptive Enterprise Search Installation, Upgrade, and Setup Guide* or the *Perceptive Workgroup Search Installation, Upgrade, and Setup Guide* and follow the instructions to preserve your indexes, sites, and settings.

The following enhancements and updates are included in this release.

Enhancements

Enterprise Search

- The product name and branding has changed from ISYS Enterprise Server to Perceptive Enterprise Search.
- The Social Explorer now allows users to search, monitor, and interact with social media from within the Enterprise Search site. An administrator can choose to enable any of the following social sources for users to search: Twitter, Facebook, LinkedIn, and Flickr.
- After significant speed optimizations, manual and scheduled index updates now occur at an enhanced rate.
- The following connectors are now available:
 - Perceptive Content (ImageNow) Federation. Unifies access between your Perceptive Software content deployment and other data sources with Enterprise Search. Full-text content search is available when you deploy Content Server.
 - Perceptive (formerly Brainware) Globalbrain Federation. Unifies access to your data repositories and your Globalbrain deployment with a single search.
 - Twitter Federation. Blends and navigates enterprise results alongside results from Twitter.
- Product certifications now include the following platforms:
 - Windows Server 2012
 - Windows 8 32-bit and 64-bit (Perceptive Enterprise Search Anywhere)
 - iOS6 (Perceptive Enterprise Search Anywhere)
 - Android 4.x (Jelly Bean) (Perceptive Enterprise Search Anywhere)
- A user can now set and save preferences to customize the search experience. Preferences include language conflation (stemming), use of the thesaurus, default sort order, results per page, and the Social Explorer. User selections in the Advanced and Menu-Assisted search are also saved and automatically populated across searches.
- An administrator can now configure web proxy settings through Global Web Proxy. The administrator can configure a connector's proxy settings at the connector or global level.
- ZIP file creation when exporting results is now faster.

- Document processing now extracts more data, such as equations and bookmark locations, from supported file types. Additionally, document filtering now provides enhanced hyperlink support, extraction of documents embedded in Microsoft OneNote, and enhanced Microsoft PowerPoint support.

Workgroup Search

- The product name and branding has changed from ISYS Workgroup to Perceptive Workgroup Search.
- Windows 8 for both 32-bit and 64-bit is now a certified platform.
- ZIP file creation when exporting results is faster.
- After significant speed optimizations, manual and scheduled index updates now occur faster.
- Document filtering now extracts more data, such as equations and bookmarks, from supported file types. Additionally, document filtering provides enhanced hyperlink support, extraction of documents embedded in Microsoft OneNote, and enhanced Microsoft PowerPoint support.

Updates

Core Search

The following core updates affect all Perceptive Search products.

- HiDef output now preserves white text on a black background when found in Microsoft Office Word documents (DOC and DOCX). (FB7756)
- The output for Microsoft Office PowerPoint PPTX files now includes embedded images. (FB7947)
- The Document Filters version 10.3 installation wizard no longer generates a UAC elevation message. (FB7921)
- Hangul version 3 is not available for this release. (FB7932)
- Bookmark links found in the table of contents of Microsoft Office Word documents (DOC and DOCX) are now preserved. (FB7934)
- HiDef output now includes all images from a Microsoft Office Word document. (FB7946)
- When running the Java sample with the BuildWindows.bat file, the "Access Violation" message no longer appears. (FB7969)
- The mapping of open errors for PDF files no longer incorrectly assigns errors in the ISYS.log file. (FB7963)
- Hit-highlighting is now present for MHTML files. (FB7968)
- Dates are now correctly formatted in Microsoft Office Excel (XLS) files when using the convertdocumenttoutf8.exe file. (FB7993)
- HiDef output is no longer missing images for OST files. (FB7958)
- HTML view now includes email attachment icons. (FB7997)
- Default index options now include timing logs in the ISYS.CFG log file. (FB7985)
- Documentation now includes usage information on the IndexWorkers option. (FB7901)

- The Enterprise Search installation wizard no longer displays "The parameter is incorrect" message in the ISYS catalog file. (FB7866)
- Allowed hits and entities are now highlighted differently in PDF files that are submitted to OCR. (FB7838)
- The "Edit Template" option found under "Template" for a Website no longer displays an HTTP 500 error. (FB7815)
- Output is now generated for an MSG file that is converted to text when the "To" or "CC" fields contain more than 100 characters. (FB8008)
- The "Last Update" field now updates after an index is updated in Server Tools. (FB7802)
- The MBOX Last Modified date is now correct. (FB8015)
- The metadata is now the same for both the DOC and DOCX file types. (FB8013)
- A scripted index no longer generates an AV error after approximately an hour of indexing. (FB8023)
- Korean characters can now be extracted from XPS files. (FB8022)
- When an email message is stored as an MSG file type and indexed using HiDef, the header now contains the email message's sent date. (FB8009)
- The contact information in the License Code activation window now lists the new telephone numbers. (FB8016)
- The handling of linked numbers and table of contents entries in Microsoft Word documents is now improved. (FB5972)
- An issue where a temporary file was sometimes orphaned when documents were processed is now resolved. (FB6624)
- Microsoft Excel reader now provides enhanced performance when processing files in HiDef mode. (FB6648)
- Track Changes comments can now be extracted from Microsoft Word documents. (FB6680)
- The handling of metadata when content contains HTML tags is now improved. (FB6684)
- Shift-jis text files are no longer misidentified as xy write. (FB6693)
- A password-protected ZIP file previously returned the generic error code 4. Now a password-protected ZIP file returns error code 5. (FB6729)
- The handling of corrupted RAR files is now improved. (FB6730)
- The handling of PDF files embedded in Office 2007 files is now improved. (FB6731)
- HTML mode now supports rotated text found in Microsoft Word documents. (FB6831)
- All supported platforms can now process a PST file that contains a MSG file (email message) with an embedded MSG file. (FB6864)
- The handling of Smart tag items in Microsoft Word documents is now enhanced. (FB6883)
- The handling of numbered paragraphs in Microsoft Word documents is now improved. (FB6901)
- The Excel HTML reader now outputs slide and date numbers in the headers and footers. (FB6950)
- Previously, document filters could not always process DWG files. Now document filters successfully extract metadata and context from DWG files. (FB6982)

- The handling of automatically generated emails that might not include content (such as calendar acceptance) is now improved. (FB7003)
- The handling of Excel files saved on the MacOS platform is now improved. (FB7079)
- The handling of non-message type MSG files is now improved. (FB7178)
- The handling of MSG files that contain multiple "body" streams is now improved. (FB7179)
- The AVI reader now inserts a paragraph break between each metadata item. (FB7189)
- The rendering of Autoshapes in Microsoft Office Word documents is now improved. (FB7209)
- The handling of large HTML files is now improved. (FB7274)
- The handling of foreign characters in the Microsoft Word 2007 reader is now improved. (FB7277)
- PST reader can now capture emails that were not previously extracted. (FB7927)
- The format detection logic of EML and MSG file types is now improved. (FB7280)
- The Microsoft Excel CSV reader now returns all discovered metadata fields. (FB7289)
- The ability for Microsoft Word reader to process form data in HTML mode is now enhanced. (FB7298)
- The handling of automatically sized tables in Microsoft Office PowerPoint 2007 is now improved. (FB7304)
- The string SHOWHIDDEN=HIDDEN is now available to format the comments of Microsoft Office documents as hidden <div>s in the HTML. (FB7305)
- The string PDFHIDETEXTLAYER=TRUE is now available to hide the text of a document gathered through OCR while still presenting the image view. (FB7310)
- The PST reader now captures emails that were not previously extracted. (FB7335)
- Uncompressed ZIP files are no longer misidentified. (FB7342)
- The 7-zip reader now returns dates in local time. (FB7351)
- The detection logic of DXF files that were previously identified as text is now improved. (FB7364)
- The processing of fractions in Microsoft Excel HTML reader is now improved. (FB7365)
- Error checking when bad parameters are passed to the API is now improved. (FB7367)
- The handling of fields within the Microsoft Word reader is now improved. (FB7370)
- The handling of corrupt PST files is now improved. (FB7372)
- Microsoft PowerPoint reader previously output text in the order it was stored. Now Microsoft PowerPoint reader outputs text in visual order. (FB7396)
- An issue in Microsoft Word HTML reader where character padding might vary by platform is now resolved. (FB7397)
- An issue in Microsoft PowerPoint reader where image handling might vary by platform is now resolved. (FB7398)
- Linux distribution compatibility is now improved. (FB7428)
- The handling of corrupt Microsoft Excel files is now improved. (FB7437)
- The identification of EML files that do not contain recipient fields is now enhanced. (FB7451)

- The handling of Microsoft Word documents that contain supplementary Unicode characters is now improved. (FB7452)
- The handling of calendar MSG files is now improved. (FB7470)
- The handling of contact MSG files is now improved. (FB7471)
- The handling of task MSG files is now improved. (FB7472)
- The handling of vCard files is now improved. (FB7473)
- An issue that occurred when sub-enumerating documents with long file names in Java is now resolved. (FB7482)
- The memory used while processing large Microsoft Office Excel files is now improved. (FB7504)
- Archive processing is now moved to the following, separate module: 7z.dll, 7z.so, 7z.dylib. (FB7535)
- RTF files now include additional metadata. (FB7537)
- The handling of Arabic RTF files is now improved. (FB7537)
- The handling of DOCM files (Macro enabled Microsoft Word documents) is now improved. (FB7552)
- HTML mode now provides improved text alignment for Microsoft Office PowerPoint 2007 files. (FB7553)
- An issue where some binder files were not processed on non-Windows platforms is now resolved. (FB7580)
- An issue where the display of Microsoft Office Project files might vary by platform is now resolved. (FB7582)
- OneNote reader now processes Unicode characters. (FB7583)
- Visio reader now displays the order of metadata consistently across platforms. (FB7584)
- The handling of Microsoft Office Excel files that contain Chart worksheets is now improved. (FB7585)
- Hangul documents are now processed on HP-UX or AIX. (FB7586)
- The output of MP3 files on AIX now matches the output of MP3 files on other platforms. (FB7587)
- The handling of corrupt TIFF files is now improved. (FB7588)
- The issue of GD warnings that streamed to the console when processing some word documents is now resolved. (FB7608)
- An issue in HTML mode with processing images that have negative margins in Microsoft Word is now resolved. (FB7609)
- A memory consumption issue when extracting some messages from PST on 64-bit non-Windows platforms is now resolved. (FB7610)
- Off-screen images in Microsoft Office Excel files are now better handled in HTML mode. (FB7612)
- Column widths and alignment in Microsoft Office Excel files are now better handled in HTML mode. (FB7632)
- Server.ProcessID is now available in ISYSScript to return the current process id. (FB7905)
- Toxic flagged documents can no longer be added to the index multiple times. (FB7949)

- Update speeds for a scripted index now prevent documents from being unnecessarily requested. (FB7955)
- The handling of when an incorrect path is passed to Init_Instance is now improved. (FB8011)
- An issue where EAM would not be considered when processing Transaction files is now resolved. (FB8012)
- CVS mode now supports Microsoft Office Excel files that include Unicode characters in the file name. (FB8059)
- An issue where a ZIP file might not be processed if it contained a zero-size file is now resolved. (FB8057)
- The following functions to calculate hashes on ISYS streams are now available (FB6635):
 - IGR_Calculate_MD5
 - IGR_Calculate_SHA1
 - CExtractor::getHashMD5
 - CExtractor::getHashSHA1
- PST reader could extract embedded MSG messages. Now it can also extract embedded EML messages by passing the string PSTMESSAGE TYPE=EML. (FB6645)
- The text and HTML modes in the Microsoft Word reader now extract the names of bookmarks. (FB6681)
- The text and HTML modes of the Microsoft Word, Excel, and PowerPoint readers now extract values from Equations. (FB6682)
- The text and HTML modes in the Microsoft Word, Excel, and PowerPoint readers now extract both the text and URL of hyperlink items. (FB6683)
- The detection of CSV files that contain little or no white space characters is now improved. (FB6863)
- Microsoft Office OneNote files now support sub-enumeration. (FB6879)
- PST now outputs email headers when EMAILMODE=AllHeaders is specified in the options string. (FB6937)
- Microsoft PowerPoint reader now offers the ability to show or hide images in the output. This is controlled by specifying SHOWHIDDEN=OFF or SHOWHIDDEN=VISIBLE. (FB7430)
- Non-Windows binaries now include a version stamp that can be displayed using "[Filename]" grep -i "\$Revision". (FB7453)

Enterprise Search

- The "Enable Metadata Refinement" option now persists after it is enabled. (FB7884)
- An issue where internal relative links were missing characters in Microsoft Office Word documents and causing invalid links is now resolved. (FB7856)
- The "Last Run" column on a Maintenance Plan now displays information after the plan is run. (FB7801)
- Rebuilding indexes through the Preferences window no longer removes the Language Unicode setting in the ISYS.cfg file. (FB7796)

- Enterprise Search can now add users from a domain with more than 25,000 users. (FB7811)
- Multiple document management system rules on a single index are now supported. (FB7783)
- An issue where the icons for results from a federated source defaulted to the first result's icon is now resolved. (FBN/A)
- Case refinements now support federated sources. (FB8014)
- An issue where the blank.zip file was missing from the web site folder has been resolved. (FB8020)
- When Menu-Assisted, Advanced, or Natural Language searching is used in the query window and the user does not select a default index, a default index is now automatically selected. (FB8001)
- The account name no longer changes after the "Index Updates" interval is modified. (FB7935)
- Enterprise Search Alerts now allows notifications for both Success and Failure messages. (FB7951)
- Synonym Rings now correctly displays Nordic characters. (FB8028)
- An issue where some URLs returned through IIS were invalid is now resolved. (FB8053)
- The processing of Excel HTML is now improved. (FB6960)
- HTML and CSV modes have better processing for corrupt Microsoft Office Excel files. (FB7285)
- The detection logic on EML for incomplete headers is now improved. (FB7326)
- Sub-enumerated messages from a SendMail file now report the correct date. (FB7400)
- The font scaling of text boxes in Microsoft Office Excel files is now improved. (FB7525)
- A warning message now appears when an administrator attempts to bind to the root of an IIS web site. (FB6025)
- The product documentation is now updated to reflect the new product name, enhancements, and updates. (FB7718)
- The handling of viewing a deleted SharePoint document is now improved. (FB7823)
- SharePoint indexer now enables a lower privileged user to index the site. (FB7824)
- Synonym Ring now supports Chinese characters. (FB7847)
- An issue where a download might fail if a file name contains Chinese characters is resolved. (FB7849)
- Metadata Refinement counts from searching multiple indexes are now correct. (FB7938)
- Web Style searches now offer improved handling of Quotes. (FB7950)
- An issue where some indexes reported "Rebuild required" after an initial index is now resolved. (FB8005)

Workgroup Search

- Selecting ALT + F and then clicking the "X" (exit) button now closes the program. (FB7977)
- The "Macros / Extract documents to zip" option previously did not retain the time it extracted a PST file, it created redundant file names, and it sometimes stored the MAPI ID in the file name. These issues are now resolved and the time to extract data from a PST file is now enhanced. (FB7976)
- An issue where exporting PST files would generate files with no extension is now resolved. (FB7975)
- The unpacking of ZIP files is now optimized. (FBN/A)
- The redirection in spider indexing is now improved. (FB7945)
- Product documentation is now updated to no longer include references to the deprecated "Browse PDF documents using Adobe Acrobat" option. (FB7880)
- The cursor focus no longer moves to the top of the results list after a results item is deleted. (FB7864)
- The "In Folder" column now sorts correctly. (FB7862)
- Changes to the registry are now written when the location of the Rlist files are changed by using the "Results Cache" option under "Maintenance". (FB7861)
- Previously, when a user clicked the Print button on the Document tab, the document was immediately printed. Now the user is prompted to select printer options before the document is printed. (FB7772)
- An issue where files assigned file names that started with "ISYS" were not indexed is now resolved. (FB7778)
- The copyright year in all script templates is now up to date. (FB7787)
- Workgroup Search would sometimes stop a copy function when a file could not be found. Now Workgroup Search completes the copy function and lists the error on the log file. (FB7786)
- Catalogue names are no longer truncated. (FB7713)
- Support for Workgroup Search in a Terminal Server environment is now enhanced. (FB7994)