

# **ABBYY FlexiCapture 9.0**

## 文档处理指南

亲爱的用户！

本指南对安装项目中文档处理期间所执行的操作进行了描述。

如果还要了解更多信息，请尝试以下方式：

- 使用其它帮助文件。使用程序菜单或单击“F1”，或单击“开始 > 程序 > ABBYY FlexiCapture 9.0 工作站 > 帮助”，打开帮助文件。

希望您喜欢使用我们的产品！

# 目录

1. 介绍.....	3
1.1. 数据捕获的用途.....	3
1.2. 数据捕获自动化.....	3
1.3. ABBYY FlexiCapture 9.0 可以处理的文档.....	4
1.4. 数据控制.....	4
1.5. ABBYY FlexiCapture 9.0 校验工作站.....	5
2. 文档处理.....	5
3. 任务.....	6
4. 组装检查.....	7
4.1. 添加图像.....	8
5. 数据校验.....	8
5.1. 组校验.....	8
5.2. 字段校验.....	9
6. 校验.....	10
7. 批处理文件完整性检查.....	11
8. 导出和导出确认.....	11
9. 处理异常.....	11
10. 键盘快捷键.....	11
10.1. 主窗口.....	11
10.2. 组校验窗口.....	13
10.3. 字段校验窗口.....	14

## 1. 介绍

### 1.1. 数据捕获的用途

现在的商业、生产和服务领域都会用到大量文档。申请表、调查问卷、发票、草稿以及其他文档，是任何一家公司的必备之物。现代信息技术突显了纸质文档的不足，大部分数据被转换为电子形式来进行存储、分析和处理。

电子文档最耗时耗力的就是数据输入。只能凭借手工输入数据，输入少量信息尚可，但如果有大量文档，就勉为其难了。由于手工输入团队难以管理，并且更改的成本与从头开始处理的成本相当，因此在如此严苛的形势下，手工输入速度不能得到马上提高。

因此，手工输入并非最佳途径。另一种更简单也更有效的途径，就是使用像 ABBYY FlexiCapture 9.0 之类的自动数据捕获系统。

### 1.2. 数据捕获自动化

ABBYY FlexiCapture 9.0 是一款用于处理结构化、半结构化 (FlexiLayout) 和非结构化文档的数据捕获软件。

自动数据捕获由以下阶段组成：

- 首先使用文档扫描仪对一系列页面进行**扫描**；
- 扫描页面被自动组装成文档；
- 自动**识别**字符；
- 将不确定识别的字符发送给操作员**检查（校验）**，
- 确认的数据最终将会**导出**至文件或数据库，文档图像将会保存至指定文件夹。这些图像可以采用图形格式或作为可搜索 PDF 文件保存。

ABBYY FlexiCapture 9.0 是一款高效的数据捕获自动解决方案，它可以轻松控制工作进度和质量。

### 1.3. *ABBYY FlexiCapture 9.0 可以处理的文档*

ABBYY FlexiCapture 9.0 是一种支持不同文档类型的数据捕获应用程序。

ABBYY FlexiCapture 9.0 可以处理的文档类型如下。

- **结构化文档**。包含专用数据字段且这些字段在各文档副本中的数量、位置和格式均保持不变的文档，即结构化文档。此类表单通常以印刷形式发行，并且要手工填写。要识别结构化表单和捕获表单数据，必须创建布局向本程序指示字段位置。布局在项目安装期间由 ABBYY FlexiCapture 9.0 创建。
- **半结构化文档**。各副本的数据字段在数量、位置和格式不同的文档，即半结构化文档或弹性文档。发票就属于这类文档，其原因是发票由不同公司开具，发票的项目和格式通常有所不同。所有发票都有账户号和支付金额，但这些项目位于文档的不同部分。ABBYY FlexiCapture 9.0 使用 FlexiLayout 来识别半结构化文档和数据捕获。FlexiLayout 在 ABBYY FlexiLayout Studio 中进行创建。有关 ABBYY FlexiLayout Studio 的信息，请参阅其帮助文件。弹性文档和半结构化文档的处理，仅创建阶段和布局匹配阶段与结构化文档不同。
- **非结构化文档**。如果您要处理包含自由表单信息的非结构化文档，如合同、信函、订单、图表，ABBYY FlexiCapture 9.0 也是理想的解决方案。非结构化文档可以自动作为结构化文档或弹性文档的补充来识别，或者通过使用 FlexiLayout，然后导出至图像文件和可搜索 PDF 文件。索引字段可自动或手动地从非结构化文档捕获（使用 FlexiLayout）。典型的非结构化文档处理方案，就是将纸质存档转换为电子形式，捕获几个基于属性搜索所需的索引字段。

### 1.4. *数据控制*

识别的数据可由校验操作员控制，或者自动控制。

操作员执行检查，如有必要，可纠正不确定识别的字符。此过程就是校验过程，分两个阶段执行。

- 第一阶段是组校验 — 对识别为相同值的字符图像分组，并显示在操作员屏幕上，以便查看和纠正突出于其余部分的字符。
- 第二阶段是字段校验 — 校验特定字段值。

系统自动执行如下数据控制操作：

- 检查页面是否正确组装成文档。
- 检查数据与指定格式是否一致。
- 检查是否符合规则。规则即项目安装时指定的文档字段的特定相关性。
- 批处理文件完整性检查。

如果在检查过程中发生错误，相应的文档（批处理文件）将被发送给操作员进行检查。

## 1.5. ABBYY FlexiCapture 9.0 校验工作站

ABBYY FlexiCapture 9.0 中检查和校验文档所使用的工作站如下：

- 数据校验工作站 – 用于组校验和字段资料校验；
- 校验工作站 – 用于所有文档检查。

## 2. 文档处理

文档将由操作员处理或者自动分几个阶段处理。所有这些阶段有严格的次序，必须先通过上一个阶段，才能开始当前阶段。处理基本上分为这些阶段：图像导入、识别、校验和导出。

每个阶段会形成一个处理任务队列。在处理期间，文档由初始队列转向最终队列 — 导出。在校验阶段，可以创建检查和校正特定错误的独立队列，以提高处理速度和处理质量。

数据导入、识别和导出是自动执行的。因此，操作员可以访问如下队列：

- 文档组装检查；
- 数据校验 — 对不确定识别的字符和格式错误进行组校验和字段校验；
- 校验 — 规则和格式错误校验（在文档窗口中）；
- 批处理文件完整性检查；
- 导出确认；
- 异常。

操作员可访问的队列由操作员角色进行定义。数据校验操作员只能访问数据校验队列。校验操作员可以访问除“导出确认”和“异常”以外的所有队列。高级校验操作员可以访问所有队列。

操作员可以访问的队列列表在项目安装期间进行配置。需要选择“校验”和“异常”队列。如果未选择“文档组装检查”和“数据校验”队列，将在“校验”队列内部执行相应处理。如果检查了批处理文件完整性，但未选择“批处理文件完整性检查”队列，有错误的批处理文件将显示在“异常”队列中。如果未选择“导出确认”队列，将在校验结束后立即导出。

在识别或导出期间发生错误的文档或批处理文件，也将被发送至“异常”队列。

### 3. 任务

文档被发送给操作员，供操作员在任务中进行处理。任务中的文档数因队列和项目设置而异：


- 如果要将多个批处理文件作为一个整体处理，每个批处理文件中的所有文档将被添加至该任务。在“文档组装检查”、“批处理文件完整性检查”和“导出确认”期间，批处理文件作为一个整体处理。如果在项目设置中进行了配置，同样也可以在“校验”队列中进行此类处理。
- 如果不需要将批处理文件作为一个整体来处理，每个任务将包含 10 个文档。如果批处理文件中的文档数不是 10 的倍数，将由奇数个文档构成一个独立的任务。因此，您会获得多个任务，每个任务有 10 个文档，也可能有 1-9 个文档。此处所述的程序适用于“数据校验”队列和（在进行了所需设置的情况下）“校验”队列中的任务。


一个任务只能由来自同一个批处理文件的文档组成。



操作员的工作是从选择队列和接收任务（图 1）开始的。



图 1. 主窗口

要获取任务，单击  获取任务。当任务来自“数据校验”队列时，系统将打开组校验或字段校验窗口。当任务来自任何其它队列时，程序将打开文档列表。

完成任务处理时，必须单击  关闭任务 来关闭任务。然后，文档（批处理文件）将被发送至下一个处理阶段。如果任务来自于“数据校验”队列，完成数据校验时该任务将自动关闭。

使用  按钮可取消任务（拒绝任务）。对于这种情况，该任务将被分解并返回给其初始队列。您也可以单击（）来推迟任务。被推迟任务的列表显示在主窗口下方（请参阅图 1）。通过双击可以打开被推迟的任务。其他操作员无法查看被推迟的任务。

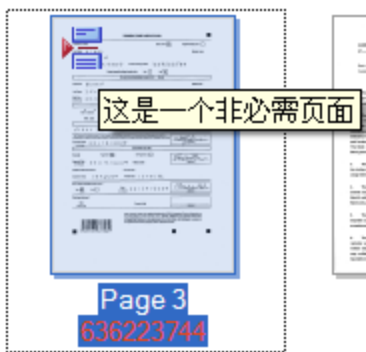
如果在处理过程中发生非预期情况，可以单击  将任务发送至“异常”队列。高级校验操作员可以将任务发送给任何队列，也可以单击此按钮将其分配给某操作员。

任务菜单命令可用于将任务中的某些文档发送至“异常”（将文档发送至异常...），将文档添加至现有“异常”（将文档添加至异常...），删除文档（从异常中删除文档...），以及查看现有“异常”列表（查看异常列表...）。“异常”列表显示每个异常的注释。在向“异常”列表添加任务或文档时，会添加注释。“异常”队列中的任务由高级校验操作员处理。

## 4. 组装检查

在分页文档中，将执行文档组装检查。页面将自动组装进文档中。如果出现任何错误，只需要纠正错误。

如果页面顺序与指定的不符，或者不同页面上的关键字段值不匹配，则该文档将标为红色，并且文档窗口中会显示相应的错误消息。



对于这种情况，请检查扫描过程中是否混合了这些页面。有时，通过更改页面顺序就可以纠正组装错误。

文档组装可以在缩略视图模式的页面中轻松检查（图 2）。您可以更改页面位置，甚至可以采用拖放法将页面从一个文档移至另一个文档。缩略图视图模式还整合了弹出提示，如图所示。

如果必须为正确的文档组装执行关键字段校验，将在每页图像的下面显示关键字段值（请参阅图 2）。如果文档中的关键字段不匹配，字段值将标为红色。

由于关键字段可能填写或识别错误，因此可能会不匹配。请检查关键字段值。如果关键字段值仍然不匹配，请检查这些页面是否属于同一个文档。如果页面顺序错误，请查找关键字段值匹配的页面，并将这些页面组装到文档中。

注：您可以缩放页面缩略图。按 Ctrl 键并上下滚动鼠标滚轮即可缩放。

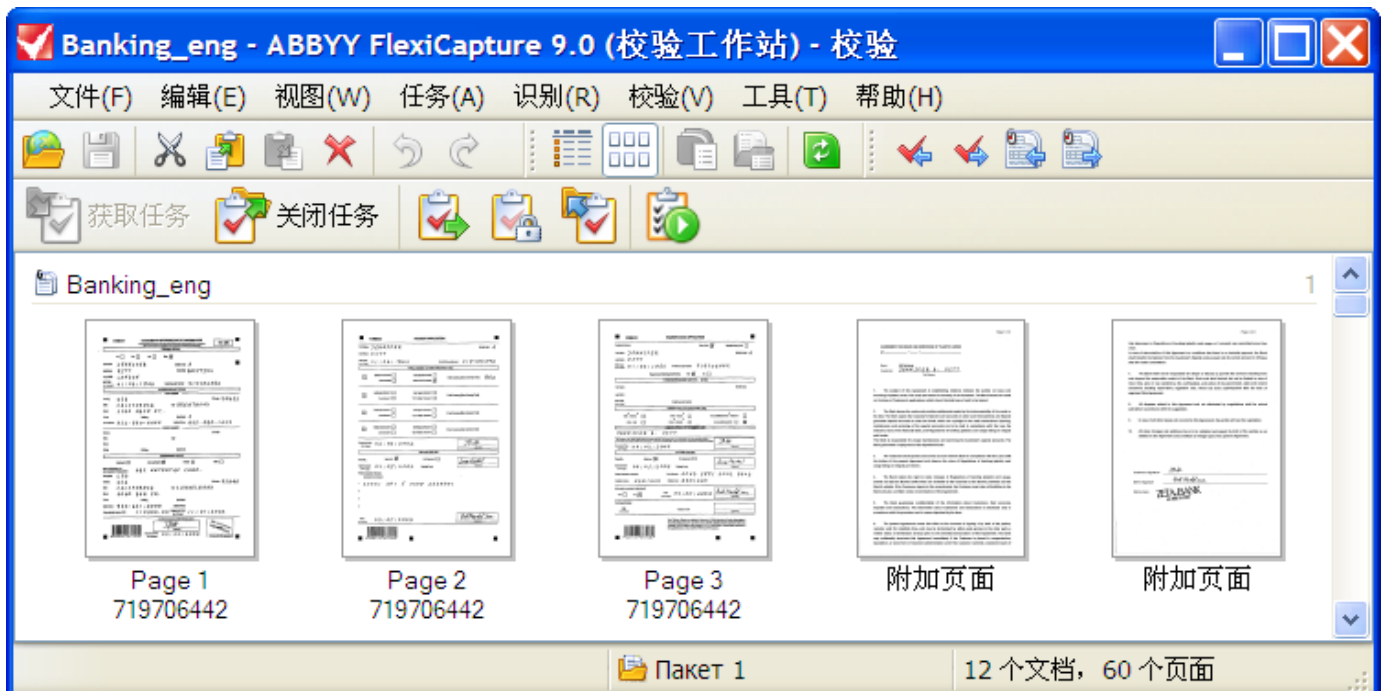


图 2. 缩略视图模式中的主窗口

完成组装错误纠正后，请关闭任务。如果错误仍然存在，系统将发出警告。

## 4.1. 添加图像

图像可以自动从热文件夹导入至系统，也可以由操作员在扫描工作站导入。在必要时，校验操作员或高级校验操作员可以扫描缺少的图像（文件 > 扫描图像...），或加载缺少的图像（文件 > 加载图像...）。这对于纠正组装错误十分有用。

# 5. 数据校验

在从“数据校验”队列获取任务时，将打开“组校验”窗口。完成“组校验”后，会打开“字段校验”窗口。任务完成时，窗口会自动关闭。

## 5.1. 组校验

**组校验**即检查识别为相同值的分组字符图像。所有被识别为完全相同的字符（例如图 3 上的数字 0）将显示在校验程序中，以用于确认正确的变体，只留下不正确或不确定的字符供下一阶段处理。

如果不确定字符的值，请执行如下操作：

1. 在选定字符的本地菜单中选择**显示字符图像**，或按 F2。将显示该字符所在字段的图像。
2. 单击校验程序窗口中的**视图 > 字段图像 > 显示字段图像**，或按 Ctrl+I。当将鼠标指针移到被校验字符时，将启用弹出字段。

如果字符识别不正确：

1. 选择与组字符不一致的字符，并指定正确值。此值将在字符图像左上角标记为绿色。
2. 如果在检查上下文后仍不能确定字符值，请在该字符上单击左键插入一个红色问号。也可以使用**更改状态**按钮更改字符状态。
3. 如果需要删除字符，可选中该字符并按 **Del**。

正确识别的字符必须通过单击**确认**予以确认。要确认显示的所有字符，单击工具栏上的**全部确认**按钮。





图 3. 数字的组校验

## 5.2. 字段校验

**字段校验**即对带有已知值区域的字段上下文中不确定识别的字符进行检查。例如，我们知道**国家**字段应采用哪些值，因此如有必要，可以很容易地纠正该字段值。

可以采用标准的文字处理器模式，例如字符插入和替换，来纠正错误识别的符号。按 **Insert** 可切换模式。

逐一检查每个字段的识别结果，进行必要的修改，然后按 **Enter** 或单击**确认字段**来确认值。

对于其值与指定数据类型不符的字段，将标以红色的 **F**，并显示一条错误消息。对于指定了规则但字段值与规则不符的字段，情况同样如此。请纠正此类字段的值。如果无法纠正，请单击**推迟字段**来推迟值的检查，以后在“校验”阶段进行纠正（文档编辑器窗口）。

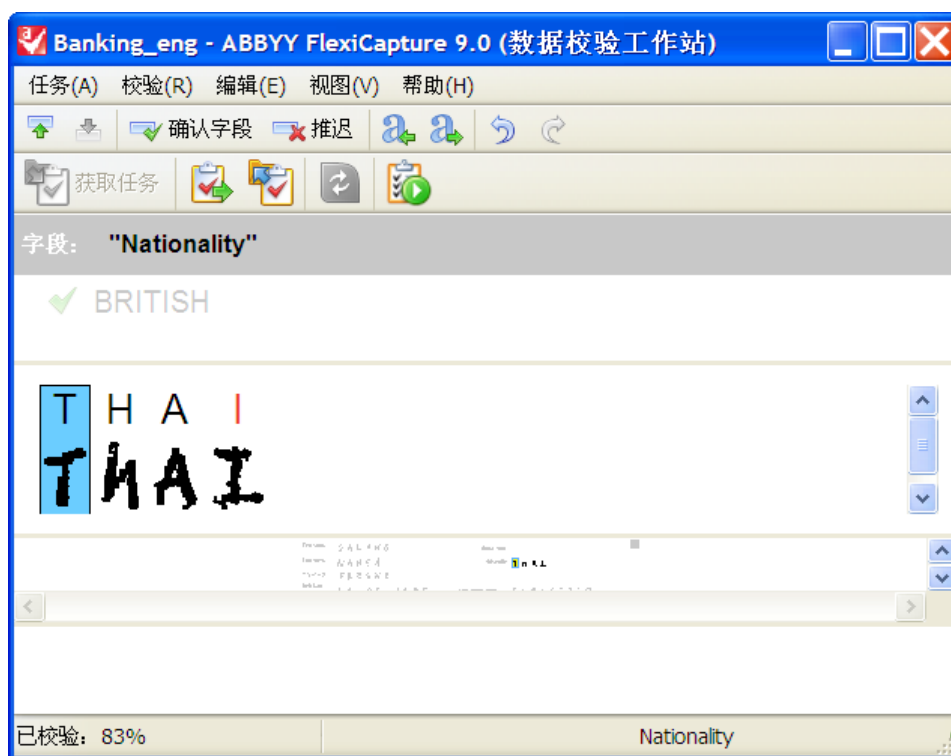




图 4. 字段校验窗口

## 6. 校验

校验期间执行的检查如下：格式检查（识别的值不符合字段格式），以及规则检查（如果在项目安装期间指定了规则，字段数据必须对应这些规则）。如果“组装检查”队列和“数据校验”队列未单独指定，这些过程将在“校验”任务中执行（请参阅“组装检查”和“数据校验”章节）。

格式错误和规则错误在文档编辑器窗口中纠正。双击页面或文档的名称可打开该窗口。这个窗口中可能会显示识别的数据或文档图像。另外，文档编辑器还整合了错误窗口。单击**布局**按钮，可自定义窗口位置。

要校验的字符标为红色，数据类型不正确和不符合规则的字段则为突出显示。使用  和  按钮，可选择上一个和下一个校验对象。您可以逐一查看组装错误、不确定识别的字符、规则错误以及其他项目。

规则问题按以下情况进行标记：黄色表示警告，红色表示错误。规则警告和错误显示在文档编辑器的单独窗口中，不符合规则的文档在文档列表中标有红色标记。

如果有某些规则不符，操作员必须检查数据校验结果。如果出现校验错误，请纠正这些错误。如果有无法纠正的填写错误，请将文档发送至“异常”队列。

纠正错误后，会自动重新检查该规则。单击**工具 > 重新核对规则**，可手动重复校验。

## 7. 批处理文件完整性检查

在导出期间，管理员所指定的脚本将执行批处理文件完整性检查。如果存在错误，该批处理文件将被发送至“批处理文件完整性检查”队列（或者，如果没有该队列，则发送至“异常”队列）。管理员在批处理文件完整性检查脚本中会指定任务注释。在生成的任务中，纠正那些破坏批处理文件完整性的错误并关闭该任务。

## 8. 导出和导出确认

“导出确认”队列适用于高级校验操作员。在接收来自该队列的任务时，请检查任务文档是否可导出。关闭任务时，将开始执行批处理文件完整性检查。如果检查成功，将自动导出文档。如果不导出文档或任务，可推迟该任务，或将该任务（或部分任务）发送至其他处理阶段或发送至“异常”。

如果完整性检查期间发生错误，系统将发出警告：“批处理文件 <批处理文件名称> 存在完整性错误。是否继续导出？”单击**是**以继续，单击**否**以取消。如果选择取消导出，该批处理文件将被发送至“批处理文件完整性检查”队列（或者，如果没有该队列，则发送至“异常”队列）。

高级校验操作员也可以执行本地导出。要执行本地导出，请在批处理文件视图模式中选择要导出的文档，并单击**文件 > 导出至...**，然后选择所需的命令。

## 9. 处理异常

在识别或导出期间发生错误的批处理文件和文档，将被发送至“异常”队列。如果启用了批处理文件完整性检查，但未为这些检查创建独立的队列，则存在完整性错误的批处理文件也将被发送至“异常”队列。校验操作员也可将处理期间发生不可预见错误的任务或文档填入“异常”队列。

只有高级校验操作员才能访问“异常”文件夹的任务。在纠正完错误时，必须将这些文档或任务发送至所需的阶段继续进行处理。

## 10. 键盘快捷键

### 10.1. 主窗口

保存	Ctrl+S
新批处理文件	Ctrl+N
加载图像	Ctrl+O
扫描图像	Ctrl+K
导入图像	Ctrl+I

将数据导出至文件	Alt+Shift+S
导出至数据库	Alt+Shift+D
撤消	Ctrl+Z
恢复	Ctrl+Y
剪切	Ctrl+X
复制	Ctrl+C
粘贴	Ctrl+V
删除	Del
全选	Ctrl+A
查找	Ctrl+F
查找下一个	F3
下一个文档	Ctrl+D
上一个文档	Ctrl+Shift+D
图像去斑	Ctrl+Alt+K
反转图像	Ctrl+Alt+V
顺时针旋转 90°	Ctrl+W
逆时针旋转 90°	Ctrl+Shift+W
批处理文件	Ctrl+B
队列	Ctrl+Q
详细信息视图	Alt+1
缩略图视图	Alt+2
全屏	F11
刷新	F5
属性	Alt+Enter
文档编辑器：图像模式	Ctrl+1
文档编辑器：数据模式	Ctrl+2
放大	Ctrl +Num+
缩小	Ctrl+Num-
缩略图大小：增大	Ctrl+Shift+Num+
缩略图大小：缩小	Ctrl+Shift+Num-
获取任务	Ctrl+G

分析	Ctrl+E
匹配文档定义	Alt+Shift+E
识别	Ctrl+R
运行校验	F7
下一个要校验的项目	F4
上一个要校验的项目	Shift+F4
下一个组装错误	F9
上一个组装错误	Shift+F9
下一个不确定字符	F8
上一个不确定字符	Shift+F8
下一个规则错误	F6
上一个规则错误	Shift+F6
更新至最新版本	Alt+Shift+U
重新分析	Ctrl+Alt+E
重新识别	Ctrl+Alt+R
重新核对规则	Ctrl+F6
检查批处理文件完整性	Ctrl+Shift+E
帮助主题	F1
退出	Alt+F4

## 10.2. 组校验窗口

全部确认	Enter
全部推迟	Ctrl+Enter
更改状态	Space
下一页	Page Down
上一页	Page Up
撤消	Ctrl+Z
恢复	Ctrl+Y
全选	Ctrl+A
显示字符图像	F2

全屏	F11
显示字段图像	Ctrl+I
字段图像：在顶部	Alt+1
字段图像：在底部	Alt+2
放大	Ctrl+Num+
缩小	Ctrl+Num-
帮助主题	F1
退出	Alt+F4

### 10.3. 字段校验窗口

确认字段	Enter
推迟	Ctrl+Enter
下一个不确定字符	F4
上一个不确定字符	Shift+F4
下一字段	Page Down
上一字段	Page Up
撤消	Ctrl+Z
恢复	Ctrl+Y
剪切	Ctrl+X
复制	Ctrl+C
粘贴	Ctrl+V
删除	Del
全部删除	Alt+Del
全选	Ctrl+A
插入分行符	Shift+Enter
合并字符	Ctrl+M
类似字段	Alt+F3
显示字符图像	F2
全屏	F11
字段数据：识别的文本	Alt+F1

字段数据：字符图像切割	Alt+F2
显示字段图像	Ctrl+I
字段图像：在顶部	Alt+1
字段图像：在底部	Alt+2
放大	Ctrl+Num+
缩小	Ctrl+Num-
帮助主题	F1
退出	Alt+F4